THE UNIVERSITY *of* EDINBURGH

Information Structure and the Prosodic Structure of English: a Probabilistic Relationship

Sasha Calhoun

Doctor of Philosophy

Linguistics and English Language

Centre for Speech Technology Research

University of Edinburgh

2006

Abstract

This work concerns how information structure is signalled prosodically in English, that is, how prosodic prominence and phrasing are used to indicate the salience and organisation of information in relation to a discourse model. It has been standardly held that information structure is primarily signalled by the distribution of pitch accents within syntax structure, as well as intonation event type. However, we argue that these claims underestimate the importance, and richness, of metrical prosodic structure and its role in signalling information structure.

We advance a new theory, that information structure is a strong constraint on the mapping of words onto metrical prosodic structure. We show that focus (*kontrast*) aligns with nuclear prominence, while other accents are not usually directly 'meaningful'. Information units (*theme/rheme*) try to align with prosodic phrases. This mapping is probabilistic, so it is also influenced by lexical and syntactic effects, as well as rhythmical constraints and other features including emphasis. Rather than being directly signalled by the prosody, the likelihood of each information structure interpretation is mediated by all these properties. We demonstrate that this theory resolves problematic facts about accent distribution in earlier accounts and makes syntactic focus projection rules unnecessary.

Previous theories have claimed that contrastive accents are marked by a categorically distinct accent type to other focal accents (e.g. L+H* v H*). We show this distinction in fact involves two separate semantic properties: contrastiveness and theme/rheme status. Contrastiveness is marked by increased prominence in general. Themes are distinguished from rhemes by relative prominence, i.e. the rheme *kontrast* aligns with nuclear prominence at the level of phrasing that includes both theme and rheme units. In a series of production and perception experiments, we directly test our theory against previous accounts, showing that the only consistent cue to the distinction between theme and rheme nuclear accents is relative pitch height. This height difference accords with our understanding of the marking of nuclear prominence: theme peaks are only lower than rheme peaks in rheme-theme order, consistent with post-nuclear lowering; in theme-rheme order, the last of equal peaks is perceived as nuclear.

The rest of the thesis involves analysis of a portion of the Switchboard corpus which we have annotated with substantial new layers of semantic (*kontrast*) and prosodic features, which are described. This work is an essentially novel approach to testing discourse semantics theories in speech. Using multiple regression analysis, we demonstrate distributional properties of the corpus consistent with our claims. Plain and nuclear accents are best distinguished by phrasal features, showing the strong constraint of phrase structure on the perception of prominence. Nuclear accents can be reliably predicted by semantic/syntactic

features, particularly *kontrast*, while other accents cannot. Plain accents can only be identified well by acoustic features, showing their appearance is linked to rhythmical and low-level semantic features. We further show that *kontrast* is not only more likely in nuclear position, but also if a word is more structurally or acoustically prominent than expected given its syntactic/information status properties. Consistent with our claim that nuclear accents are distinctive, we show that pre-, post- and nuclear accents have different acoustic profiles; and that the acoustic correlates of increased prominence vary by accent type, i.e. pre-nuclear or nuclear. Finally, we demonstrate the efficacy of our theory compared to previous accounts using examples from the corpus.

Acknowledgements

First of all I'd like to thank my supervisors, Bob Ladd and Mark Steedman, for all their guidance, patience and care over the past years. Bob has been a constant source of inspiration, encouragement and kindness. He has been pivotal to the development of this work on both a practical and theoretical level, not least because his own work, particularly his 1996 book, proved so central to the ideas presented here. Thanks also to Mark, who has always given his unreserved support, even when I went in unexpected directions. He offered challenging comments and interesting suggestions along the way, starting with the ideas in his work, which provided the starting point for my research. Many thanks are due to Scottish Enterprise, whose generous support through the Edinburgh-Stanford Link grants *Sounds of Discourse* and *Synthesis: Integrated Models and Tools for Fine-Grained Prosody in Discourse* made this work possible.

Thanks to the people at CSTR for all their help, particularly in their responses to my rather idiosyncratic presentations at early meetings. Also to the Prosody and Information Structure discussion group for their time and ideas over the winter of 2002/3, particularly Bettina Braun, and later Dominika Oliver, who offered valuable encouragement and suggestions. Thanks to Cassie Mayo for help with SPSS and other practical problems, Bert Remijsen for his knowledge of Praat, and computer support in both Informatics and Linguistics for resolving many technical issues.

I wish to acknowledge Paul Warren and Shari Speer for their kindness in providing me with the recordings from their SPOT experiments. Although the results using their corpus did not prove conclusive for my own research, they helped in the formulation of the corpus work presented here and their generosity was greatly appreciated.

A big thanks to David Beaver for arranging my trip to Stanford in 2005, and to him, Dan Jurafsky and the members of the Prosody Discussion group for the enthusiasm and advice which they offered me in developing my, as yet, nebulous ideas while I was there. The trip was vital for the consolidation of my theories and my confidence. I am further grateful to Dan and David for the practical backing since then needed to produce the corpus used here. Thanks are also due to Jason Brenier and Florian Jaeger for many beneficial discussions, and to Jason for all the help and encouragement since.

The work presented here would not have been possible without the corpus development carried out by Jean Carletta and her team, particularly Neil Mayo and Jonathan Kilgour, and I am obliged to them. A big thanks to Shipra Dingare as well for her work on the tedious job of transcript alignment. Many thanks to Joanna Keating and Joseph Arko who carried out the kontrast annotation, and Hannele Nicholson, who did the bulk of the prosody annotation. Also, I am very grateful to Mari Ostendorf and her colleagues, who generously agreed to include their prosody annotations in our project.

Finally, many thanks to all the people along the road who brought a smile to my face when it all threatened to get the better of me. In particular, thanks to Bea, Ben, Anna and Shipra, who've been fantastic friends throughout, and to Billy who was such a good mate at the beginning. A huge thanks to my parents, whose boundless love and generosity lie at the heart of what I can do, and who never complain that I'm away for so long. Last, and furthest from the least, to Raj, who has loved me, inspired me and believed in me. Thank you for always being such a wonderful person to be near.

Table of Contents

1 Introduction **1**

1.1 Information Structure and the Prosodic Structure of English 3

1.2 Overview 5

2 Information Structure and Intonational Meaning **8**

2.1 Lexical and Syntactic Effects 9

2.2 Information Structure 11

2.2.1 Focus 12

2.2.2 Givenness 21

2.2.3 Theme/Rheme Structure and Contrast 29

2.3 Intonational Meaning 34

2.3.1 Richer Information Structure 35

2.3.2 Illocutionary Force 38

2.3.3 Affective Connotations and Emotion 40

2.4 Overview and the Way Forward 45

3 How Prosody Conveys Information Structure **49**

3.1 Prosodic Concepts 51

3.1.1 Phrasing 51

3.1.2 Prominence 54

3.1.3 Intonation Events 59

3.1.4 Global Pitch Variation 66

3.2 The Relationship Between Prosodic and Information Structure 68

3.2.1 Association of Focus and Nuclear Prominence 69

3.2.2 Pre- and Post- Nuclear Accents and Ambiguity 73

3.2.3 Theme/Rheme Status by Relative Prominence 77

3.2.4 Emphatic Accents and *Restricted* Kontrast 82

3.2.5 Interacting Phonetic Cues 85

3.3 The nature of Prosodic Units 87

3.3.1 A Constraint-Based Approach to Information Structure Interpretation 88
3.3.2 Intonational Tunes and Categorical Perception 93
3.4 Summary and the Next Steps 96

4 Searching for Contrastive Accents **99**
4.1 Previous Experimental Work on Contrastive Focus 100
4.1.1 Phonetic Characteristics of L+H* and H* 100
4.1.2 Interpretative Differences between L+H* and H* 103
4.1.3 Interpretation and Realisation of 'Contrastive' Accents 106
4.2 Experiments on the Nature of Theme and Rheme Accents 110
4.2.1 Experiment 1: Production Experiment 110
4.2.2 Experiment 2: Perception Experiment 115
4.2.3 Experiment 3: Relative Height 122
4.2.4 Experiment 4: Production Experiment 126
4.3 Summary and General Discussion 134

5 *Switchboard* in NXT: A Data Set for Model Development **138**
5.1 The *Switchboard* Corpus in NXT 139
5.2 Discourse Semantic Features 139
5.3 Syntactic Features 142
5.4 Prosody Annotation 143
5.4.1 Related Work 143
5.4.2 Annotation Scheme 145
5.4.3 Annotation Process and Annotator Agreement 147
5.4.4 Distribution of Prosodic Types 149
5.5 Prosodic and Acoustic Features 150
5.6 Kontrast Annotation 152
5.6.1 Related Work 152
5.6.2 Annotation Scheme 153
5.6.3 Annotation and Annotator Agreement 159
5.6.4 Distribution of Kontrast Types 163
5.6.5 Kontrast Features 163
5.7 Theme and Rheme Annotation 163

6 Predicting Prosodic and Information Structure **165**
6.0.1 Related Work 166
6.0.2 Classifiers 168

6.0.3 Feature Type Groupings 171
6.1 Phrase Breaks 172
6.1.1 Aim and Method 173
6.1.2 Results and Discussion 174
6.2 Accents 179
6.2.1 A1: Nuclear Accents are Structural 180
6.2.2 A2: Nuclear Accents are Meaningful 184
6.2.3 A3: Accent and Nuclear Accent Prediction 189
6.3 Kontrast and Prominence 197
6.3.1 K1: Kontrasts are More Prominent than Expected 198
6.3.2 K2: Kontrastive Accents Raised 203
6.4 Theme/Rheme Properties 213
6.4.1 Aim and Method 214
6.4.2 Results and Discussion 216
6.5 General Discussion 221

7 Illustrations from a Dialogue **227**
7.1 Givenness 228
7.2 Focus Projection and Pre-Nuclear Accents 235
7.3 *Restricted* Kontrast and Theme/Rheme Structure 244
7.4 Connotations of Theme Accents 251

8 Conclusion **259**
8.1 Theoretical Claims and Experimental Results 259
8.2 Contribution and Future Directions 262
8.3 Applications 266

A Stimuli for Experiment 1 **268**
A.1 Block 1 268
A.2 Block 2 269
A.3 Block 3 270
A.4 Block 4 271

B Stimuli for Experiment 4 **273**
B.1 Block 1 273
B.2 Block 2 274
B.3 Block 3 275
B.4 Block 4 276

C Existing Corpus Annotations 277
C.1 Penn Treebank POS and Syntax . . . 277
C.2 Dialog Acts . . . 281
C.3 Information Status . . . 281
C.4 Phone and Syllable Alignment . . . 284

D Kontrast Decision Tree 285

E Features Used in Chapter 6 288
E.1 Glossary of Features . . . 288
E.1.1 Discourse Semantic . . . 288
E.1.2 Syntactic . . . 290
E.1.3 Phrasal . . . 290
E.1.4 Accentual . . . 292
E.1.5 Word level Acoustic . . . 293
E.2 Features Used in Each Model . . . 294
E.2.1 Phrase Break Prediction [DV: is_break] . . . 294
E.2.2 Plain v Nuclear Accent Prediction [DV: is_nuc (noaccq excl)] . . . 295
E.2.3 Plain v No Accent Prediction [DV: is_accq (nuc excl)] . . . 296
E.2.4 Nuclear v No Accent Prediction [DV: is_nuc (accq excl)] . . . 297
E.2.5 Accent Prediction [DV: is_accq] . . . 298
E.2.6 Accent Group Prediction [DV: accq_gp] . . . 299

F Full Result Tables from Chapter 6 300
F.1 Parameter Estimates for P1 . . . 300
F.2 Parameter Estimates for A1 . . . 302
F.3 Parameter Estimates for A2 . . . 303
F.4 Parameter Estimates for A3 . . . 305
F.5 Parameter Estimates for K1 . . . 307
F.6 Multivariate Test Results for K2 . . . 309

G Further Examples from Chapter 7 311

Bibliography 312

List of Figures

2.1 Marking of Contrast 32
2.2 Büring's (2003) QUD Analysis 33

3.1 Integration of Function Words into a Prosodic Word 51
3.2 Metrical Structure of Two Prosodic Words 54
3.3 Stress Clash 55
3.4 Phrase Level Metrical Structure 56
3.5 Examples of ToBI Contours 60
3.6 ToBI Finite State Network 61
3.7 Results of Pierrehumbert's (1980) *Anna/Manny* Experiment 65
3.8 Types of Pitch Span Variation 67
3.9 Pitch Declination over Multiple Phrases 68
3.10 Relative Metrical Strength and Focus Position 70
3.11 Pre-Nuclear Kontrast 74
3.12 Relative Prominence of Themes and Rhemes 76
3.13 Ladd's (1988) *And/But* Experiment Results 81
3.14 Perceived Relative Prominence of Two Peaks, Ladd et al. (1994) 83
3.15 Interacting Constraints on Prosodic and Information Structure 90

4.1 Similarity of L+H* and H* 101
4.2 Overlap in Phonetic Characteristics of ToBI Accents 102
4.3 Pitch accent shape variations in Experiment 2 116
4.4 Hypothesis 1 Results 118
4.5 Relative Prominence of Themes and Rhemes 120
4.6 Contrastive Comparison in Experiment 3 122
4.7 Examples of Full, Weak and Deaccented Themes 124
4.8 Relative Theme/Rheme Height in Experiment 4 130
4.9 Paired and Contrastive Comparisons in Experiment 4 130
4.10 Metrical Structures in Experiment 4 133

5.1 Overview of Switchboard Corpus in NXT 140
5.2 Praat Labelling Tool 145
5.3 PN and N Accents 147
5.4 Nuclear Accent Ambiguity 149
5.5 NXT Kontrast Annotation Tool 154

6.1 Example CART Tree 170
6.2 Acoustic Features by Accent Status 207
6.3 Acoustic Features of Pre-Nuclear Accents by Kontrast Status 210
6.4 Acoustic Features of Nuclear Accents by Kontrast Status 212
6.5 Acoustic Features of Theme/Rheme Status by Place 216
6.6 Peak Features of Theme/Rheme Accents 218
6.7 Acoustic Features of Theme/Rheme Accents 218
6.8 Shape of Theme/Rheme Accents 220

7.1 Acoustic Profile of (7.1) 229
7.2 Tree Structure for (7.1) 230
7.3 Acoustic Profile of (7.2) 232
7.4 Acoustic Profile of (7.5) 234
7.5 Acoustic Profile of (7.6) 236
7.6 Syntax and Metrical Tree of (7.6) 237
7.7 Acoustic Profile of (7.10) 240
7.8 Acoustic Profile of (7.13) 241
7.9 Acoustic Profile of (7.15) 242
7.10 Relative Prominence of Theme/Rheme Accents 244
7.11 Acoustic Profile of (7.20) 245
7.12 Acoustic Profile of (7.30) 248
7.13 Acoustic Profile of (7.35) 250
7.14 Acoustic Profile of (7.45) 255
7.15 Acoustic Profiles of Theme/Rheme Accents 258

List of Tables

2.1 Baumann & Grice's (2005) Scale of Accessibility and Accent Type 27
2.2 Baumann's (2005) Information Status and Accentual Marking 27
2.3 Meaning of Pitch Accents re Steedman (2006*b*) 35
2.4 Meaning of Boundary Tones re Steedman (2006*b*) 35

4.1 Pitch Accents by Information Structure, Hedberg & Sosa (2001) 104
4.2 Results from Experiment 1 113
4.3 Boundary Type by Information Status 114
4.4 Full and Weak Accents in Experiment 3 123
4.5 Scaling of H in Experiment 3 126
4.6 Results from Experiment 4: Clauses Separately 128
4.7 Results from Experiment 4: Clauses Combined 129
4.8 Scaling of H in Experiment 4 131

5.1 Annotator Agreement on Prosody 148
5.2 Prosody Event Distribution 150
5.3 Annotator Agreement on Kontrast Type 159
5.4 Reliability of Different Kontrast Types 160
5.5 Frequency of Kontrast Types 162

6.1 Phrase Break Prediction 175
6.2 Phrase Break Prediction Features: Increasing 176
6.3 Phrase Break Prediction Features: Decreasing 177
6.4 Accent v Nuclear Accent Prediction 181
6.5 Accent v Nuclear Accent Prediction Features 183
6.6 Accent and Nuclear Accent Prediction 185
6.7 Plain Accent Prediction Features (accents only) 187
6.8 Nuclear Accent Prediction Features (accents only) 188
6.9 Accent Prediction 191
6.10 Accent Prediction Features 192

6.11 Accent Group Prediction 193
6.12 Plain Accent Prediction Features (all words) 195
6.13 Nuclear Accent Prediction Features (all words) 196
6.14 Kontrast Prediction 200
6.15 Kontrast Prediction Features 201
6.16 Kontrast Prediction Features (nuclear accents) 202
6.17 Acoustic Features of Accent Status 206
6.18 Acoustic Features of Pre-Nuclear Kontrasts 209
6.19 Acoustic Features of Nuclear Kontrasts 211
6.20 Acoustic Features of Theme/Rheme Accents 215
6.21 Peak Features of Theme/Rheme Accents 217
6.22 Acoustic Differences between Theme/Rheme Nuclear Accents 217
6.23 Acoustic Features of Theme/Rheme Accents by Order 219
6.24 L and H Features of Theme/Rheme Accents by Order 220

C.1 NXT Treebank POS Tags 278
C.2 NXT Treebank Phrase Level Tags 279
C.3 NXT Treebank Function Tags 280
C.4 Shriberg et al.'s (1998) Dialog Act Types 282
C.5 Nissim et al.'s (2004) Old Subtypes 283
C.6 Nissim et al.'s (2004) Mediated Subtypes 283

F.1 Phrase Prediction Parameter Estimates 300
F.2 Plain v Nuclear Accent Parameter Estimates 302
F.3 Plain Accent Parameter Estimates (accents) 303
F.4 Nuclear Accent Parameter Estimates (accents) 304
F.5 Accent Type Parameter Estimates 305
F.6 Kontrast Parameter Estimates 307
F.7 Nuclear Kontrast Parameter Estimates 308
F.8 Accent Status Multivariate Tests 309
F.9 Pre-Nuclear Accent Multivariate Tests 309
F.10 Nuclear Accent Multivariate Tests 310
F.11 Nuclear Accent Peak Multivariate Tests 310

Chapter 1

Introduction

Our subject is the relationship between information structure and prosodic structure in English. This area is complex, and therefore engaging, because it touches upon so many areas of linguistic enquiry. In the course of its investigation, one is drawn into debates about the distributional properties of lexical items, the syntax/semantics interface and parsing, the stochastic versus symbolic nature of meaning composition, the role of implicature in discourse semantics and the division between language and paralanguage; as well the nature of phonological systems, the inventory of prosodic constructs in English, and detailed arguments about the interpretation of different phonetic signals. Information structure describes the salience and organisation of information in relation to a discourse. It has also been argued to regulate lexical selection and syntactic structure, and maybe even to define syntactic parsing. On the other hand, it is closely linked to the pragmatic interpretation of utterances, and has been claimed to signal, through implicature, illocutionary force and affective connotations. Prosody describes the phrasing, prominence and melody of speech. It has been shown to signal 'meaning' on almost every level of linguistic interpretation, from lexical distinctions and syntax structure, information and discourse structure, to emotive content. Studying the connection between these two structures is interesting because, in English at least, one of the principal cues to information structure is prosody. Therefore, in looking at this relationship, we get a window into the whole language system.

Unfortunately, most of the work on the formal semantics of prosody does not take account of the full richness of prosody as a phonological system, and the implications this has for the prosodic signals of information structure. Much of the work looks at the prosodic correlates of information structure, particularly focus, in isolation; and may be compromised by the interacting effect of other layers of meaning on these correlates. Further, many of these accounts take a rather simplistic view of the prosodic constructs involved, often just the distribution of pitch accents within a largely linear phrasal structure. The works that do

take a wider view of prosody, particularly the role of pitch accent and boundary tone type, are linked to empirically disputed taxonomies of tonal event types, and have been criticised for not being sufficiently generalisable, or indeed verifiable. On the other hand, there is considerable debate within phonetics and phonology on the correct interpretation of certain phonetic signals. In particular, there is disagreement as to the scope of prominence and phrasing structure, and the classification of intonational events. Much of this work either considers the phonetic cues to prosody completely divorced from meaning, or considers association with meaning to be a secondary purpose in prosodic description. We will see that this disassociation may have led to a misconstrual of the 'division of labour' between prosodic structure and intonation, as well as assumptions about the interpretation of properties of the pitch contour which may be misguided given the breadth of meanings conveyed by these phonetic cues. We will argue that one's view of the information structural facts to be explained is crucially dependent on one's understanding of the prosodic system. Likewise, the emphasis in description, and therefore to some extent the conception of the prosodic system as a whole, depends on the meaningful signals being considered.

In this work we look at the relationship between prosody and information structure through a variety of lenses. Given the considerations just presented, we wished to take as wide a view of both as possible, in order to capture the interacting effects of each, while keeping within a feasible research project. We begin by reviewing the literature on the prosodic signals of information structure, as well as the literature on prosody itself, to assess whether the problematic cases for standard theories of the former arise from misguided assumptions about the latter. This allows us to develop our theory in which we claim that information structure is a strong constraint on the probabilistic mapping of words onto prosodic structure. We test one aspect of this theory, the prosodic signals of *theme* and *rheme*, using established experimental methods which, unlike many such studies, look directly at the effect of variation in phonetic cues on the perception of meaning. In the remainder of the thesis, we test the predictions of our theory using a corpus of unrestricted, spontaneous speech, the *Switchboard* corpus (Godfrey, Holliman & McDaniel 1992) in N(ite) X(ML) T(echnology) format (Carletta, Dingare, Nissim & Nikitina 2004, Nissim, Dingare, Carletta & Steedman 2004, Calhoun, Nissim, Steedman & Brenier 2005), which has been annotated with a variety of semantic and prosodic features. The use of corpus-based methodology is not common in this area of research, however, we thought it was worthwhile for a number of reasons. Firstly, much of the semantic literature reported is rightly criticised for being largely based on a limited number of examples thought up by the researchers involved. We wished to test whether the predictions of our theory would hold on a wide variety of naturally occurring language. Most importantly, the very nature of the system we were looking at, i.e.

the interaction of multiple factors on both sides, is strongly suggestive of a probabilistic relationship. We wished to see if techniques commonly used in the computational linguistics field, where the probabilistic nature of language processing is widely accepted, could be used to test the predictions of discourse semantic theories. Lastly, although we have not had the time to explore this in the current work, it is hoped our findings will be able to feed back into these computational applications, particularly in improving speech synthesis.

The central question we address in this work, therefore, is how prosody signals information structure. In order to answer this, we must look at what the relevant information structure concepts are, and how they are said to be conveyed prosodically. We also look at the other 'meanings' which are said to be conveyed prosodically, and how these interact with information structure. We discuss current theories about the nature of prosody, and how these are relevant to discourse semantic theories. In the course of answering this question, we reflect upon a number of related issues, including: what the phonetic correlates of the prosodic signals of information structural categories in English are; to what extent discourse semantics can inform debate about the nature of prosodic frameworks; and the methodologies used to investigate this relationship, including introspective examples, phonetic experiments and corpus analysis.

The rest of this work presents our responses to these questions. We begin by formally defining the scope of the project, and then give an overview of the chapters to come.

1.1 Information Structure and the Prosodic Structure of English

As the title suggests, this work is about the relationship between information structure and the prosodic structure of English. We will introduce each of these in detail over the next two chapters, however, here we briefly define what we mean by information structure and prosody, as well as the scope of the project in relation to its subject language, English.

Information structure describes the salience and organisation of information in relation to a discourse. More precisely, Kruijff-Korbayová & Steedman (2003, p. 250) define information structure as:

> comprising the utterance-internal structural and semantic properties reflecting the relation of an utterance to the discourse context, in terms of the discourse status of its content, the actual and attributed attentional states of the discourse participants, and the participants' prior and changing attitudes (knowledge, beliefs, intentions, expectations, etc.)

The 'discourse' that information structure is described in relation to is characterised

broadly, as can be seen from this definition. Like Kruijff-Korbayová & Steedman (2003), we take a 'discourse' to be any "coherent multi-utterance dialogue or monologue text", though in this work we will concentrate on spoken dialogues usually involving only two participants. In the course of a discourse, we take participants to be building a *discourse model* of the set of propositions which they take to be mutually believed. The role of information structure, then, is to encode how each new utterance relates to, alters or updates the existing discourse model. Although we do not look at this in detail, we follow Kruijff-Korbayová & Steedman (2003) in assuming that the discourse model is affected not only by the actual utterances in the discourse, but also by the participants' existing shared knowledge, attitudes, gestures, etc. As will be set out more fully in Chapter 2, we consider information structure to be defined on two dimensions, broadly relating to the 'organisation' and 'salience' of information. The first, the division into *theme* and *rheme*, distinguishes the parts of the utterance that "relate to the discourse purpose, and the part that advances the discourse" respectively. The second, the division between the *kontrast*, i.e. the parts "which contribute to distinguishing [the] actual content from alternatives the context makes available", and the parts that do not (*the background*) (Kruijff-Korbayová & Steedman 2003, p. 251).

The term *prosody* is often used quite loosely to describe 'supra-segmental' features of speech (see review in Ladd 1996, ch. 1). However, here we define it to mean the phonological system which describes the structural organisation, rhythm and tune of speech (which Ladd (1996, ch. 2) takes to be part of 'intonational phonology'). This definition subsumes Shattuck-Hufnagel & Turk's (1996, p. 196) and Beckman's (1996, p. 19) definition of *prosody* as the hierarchical structure of phonological constituents and prominence, and Beckman's (1996, p.16) definition of *intonation*, which describes the pitch contour in terms of a string of phonological tonal events (see also Pierrehumbert 1980, Beckman & Pierrehumbert 1986). One of the main concerns in this thesis is to compare the usefulness of each of these in explaining the discourse effects we are interested in. Therefore, we consider them both to be components of prosody. We use the term *prosodic structure*, i.e. of phrasing and prominence, to refer to the former, and *intonation* to refer to the latter. To describe *intonation* we largely use the T(ones) and B(reak) I(ndices) system (Silverman, Beckman, Ostendorf, Wightman, Price, Pierrehumbert & Hirschberg 1992, Beckman & Hirschberg 1999), which will be introduced in Chapter 3, subject to the reservations expressed there. As we will see, the description of *prosodic structure* will be central to the theory laid out in this thesis.

The principal phonetic correlates of prosody are pitch, length and loudness, with further effects of segmental quality and reduction (Shattuck-Hufnagel & Turk (1996), Ladd (1996, ch. 1), Warren (1999)). We deliberately refer to these as 'correlates' as the prosodic

system is phonological, therefore we do not expect the relationship between prosodic structure/intonation and phonetic signals to be direct (we return to this point in Chapter 3, but see Ladd 1996, Warren 1999, Beckman, Hirschberg & Shattuck-Hufnagel 2005). In order to experimentally measure the phonetic effects of different prosodic signals, we need to rely on the acoustic correlates of these phonetic effects, i.e. $f0$ (Hz), duration (sec) and intensity (dB). However, once more this relationship is not direct, and, as we show in Chapter 5, needs to be approximated using various normalisation procedures (see also discussion in Warren 1999).

Finally, in this work we only wish to make claims about the relationship between information structure and prosodic structure in English, although we occasionally draw on evidence from other languages where relevant. We would presume, without argument, that the basic information structural properties described exist across languages, although in other languages these properties may not be primarily signalled by prosody (e.g. see Vallduví & Vilkuna 1998). Similarly, prosody across languages involves variations in prominence, phrasing and tonal events, although these constructs can be used for quite different linguistic purposes in different languages (see Ladd 1996, ch. 4). Our argument about the relationship between these structures holds only for English. We decided to restrict our scope thus because the area is already complex, and most of the previous theoretical work primarily describes English. Further, some of the prosodic effects we were initially most interested in, e.g. the signalling of themehood by tonal pitch accent type, differ in even closely related languages, and the author wanted to be able to draw on her own native speaker intuitions in carrying out this research. The solution which is presented here is suggestive of a much more broadly applicable relationship, however. The exploration of this in other languages will have to await future research. Further, we have not closely examined variation in the prosodic realisation of the relevant phenomena between dialects of English. Our experimental work used speech from a 'standard' Scottish English speaker, a 'standard' American English speaker, the perceptual judgements of a variety of British and American English speakers and a corpus including a wide variety of American English speakers. It is the intuition of the author that the claims made hold true for most varieties of British, American and Antipodean English. However, complications may arise with some 'non-standard' varieties (e.g. see discussion in Ladd 1996, ch. 4).

1.2 Overview

The basic structure of the thesis is as follows: in Chapter 2, we set out the key phenomena in the relationship between prosody and information structure which need to be explained. We introduce the basic constructs of information structure, and how they are standardly claimed

to be signalled prosodically. We review difficult cases for standard theories, and relate these difficulties to assumptions made about the nature of prosody which we will show to be misguided. We also set out evidence for the influence of lower-level factors on the prosodic signals which also convey information structure, and suggest these factors need to be better integrated into the standard accounts. Finally, we consider proposals claiming that the tonal events signalling information structure lead to higher level illocutionary and affective connotations. We submit that the validity of these proposals, and therefore the strength of the claim that tonal events signal information structure at all, is directly affected by how well they can account for the range of attested prosodic signals of these 'meanings'. In Chapter 3, we advance a quite different explanation for these information structural phenomena within the Autosegmental-Metrical prosodic framework. We set out the basic properties of the framework, i.e. phrasing, prominence and an intonational tune comprised of tonal events. We present arguments as to why prominence/phrasing structure is recursive, which is crucial to our case. Using this framework, we lay out our theory of how prosody signals information structure. The central claim is that information structure forms a strong probabilistic constraint on the mapping of the segmental string onto metrical prosodic structure. We show that when the full expressive power of metrical prosodic structure, along with the probabilistic nature of the word/prosody mapping, is taken into account; our key phenomena, including many of the problematic cases for earlier theories, are straight-forwardly explained. We end with a discussion of the underlying nature of prosodic units, including further arguments for the probabilistic processing of prosody and an analysis of how meaning is conveyed by tonal events.

In Chapter 4, we look at the issue of whether there is a special 'contrastive' pitch accent type (e.g. L+H*), as opposed to a non-contrastive accent (e.g. H*). Through a review of the experimental work on this issue, we show that this question can be decomposed into whether there is a distinct accent marking 'contrast' (*restricted* kontrast), which we had shown to be correlated with increased prominence, and whether there is a distinct accent marking themes (as opposed to rhemes). We conduct a series of production and a perception experiments which show that, although there are a number of subtle accent shape differences, the only consistent factor separating the pitch contours of themes from rhemes in contrastive contexts is relative pitch accent height, i.e. themes are lower. We argue this shows themes are less relatively prominent than rhemes at the level of phrasing that includes both units. This strengthens our general claim that the primary signal of information structure is prosodic structure, not intonational tune.

In the rest of the thesis, we test broader predictions of our theory using a small subset of the NXT Switchboard corpus. In Chapter 5, we introduce the corpus, and describe the syn-

tactic and discourse semantic features which were extracted from its existing annotations. We then detail the substantial new layers of both prosodic and kontrast annotation which we have added to the corpus, including a description of the annotation standards and annotator agreement, as well as the acoustic features which were derived from the prosody annotation. In Chapter 6, using a series of multiple regression and CART models, we show how the distributional properties of the corpus are consistent with the predictions of our theory. In particular, we show how the perception of plain and nuclear accents are constrained by phrasing, but not the other way around. We show that nuclear accents can be reliably predicted by semantic/syntactic features, especially kontrast, while other accents cannot; consistent with the claim they are directly 'meaningful'. We show that kontrast is more likely if a word is more prominent than expected given its properties. Finally, we show that pre-, post- and nuclear accents have distinct acoustic profiles, but that this is not necessary for their perception. In Chapter 7, we demonstrate more fine-grained predictions of our theory using selected examples from the corpus. Specifically, we show how givenness, focus projection, *restricted* kontrast and theme/rheme status are signalled by prominence and phrasing; and how our theory can explain examples which standard theories would not be able to. We end by considering where the intuition that themes are signalled by pitch accent type may have come from, and suggest that it comes out of the higher-level meanings correlated with themehood. Finally, in Chapter 8, we look at the possible implications of this work for related research questions, and how these findings could be used in computational applications.

Chapter 2

Information Structure and Intonational Meaning

If information structure describes the salience and organisation of information in relation to a discourse, then prosodic prominence and phrasing are intuitively central to its conveyance. From this intuition has grown an extensive body of literature trying to formalise what information structure is, and to show clearly how it is signalled prosodically. The present work aims to add to this enterprise. In this chapter, we will see that there are still many uncertainties in the description of information structure and inconsistencies about how it is claimed to be marked prosodically. We will show that many of these arise from assumptions about the nature of prosody which may not be a true reflection of its expressive power. It can appear that one's theoretical description of the semantic facts is directly affected by one's understanding of the phonetic reality.

The other thread of argument in this chapter is to evaluate to what extent information structure impacts upon and is impacted upon by prosodic signals of other levels of intonational meaning. We lay out lower-level constraints on prominence and phrasing, and consider how these affect the relationship between prosodic and information structure. Further, we review proposals claiming that information structure is signalled by intonational tune, leading to higher-level illocutionary and affective connotations. We will see that the validity of these proposals is directly affected by evidence as to how independent the meanings of tonal events comprising the intonational tune are, and by evidence of unrelated phonetic cues to these connotations. This is turn influences our notions about the prosodic cues relevant to information structure itself, that is, whether intonational tune is used to signal information structure at all, and in turn how relevant information structure is to signalling higher-level 'meanings'. We end by briefly speculating on the consequences of this discussion for our conception about the nature of prosody itself and the general approach to studying prosody,

which will be developed in the next chapter.

In this chapter, therefore, we introduce the basic constructs of information structure, and how they are standardly claimed to be signalled prosodically.[1] We lay out problematic cases for standard theories, and suggest how these difficulties come out of misguided assumptions about the nature of prosody. In the last part of the chapter, we review evidence about the relevance of information structure to signalling intonational meaning in general, particularly illocutionary and affective connotations. We relate this back to our central question about the prosodic constructs used to signal information structure. To begin, we briefly outline lower level constraints, lexical status and syntactic structure, which affect prosodic prominence and phrasing. We will see this is relevant to our claims about how information structure is signalled.

2.1 Lexical and Syntactic Effects

The principal concern in this thesis is the relationship between information structure, and prosodic prominence and phrasing. However, there seems to be reasonable evidence that properties of the segmental string on which information structure is built, i.e. lexical items and syntax, have an independent effect on both prominence and phrasing. While the information structure theories described below usually side-step these effects by dealing with isolated language examples, we believe they are an important part of the story in describing prosodic patterns over language as a whole.

Within lexical items, certain syllables are perceived as more prominent than others, i.e. they are *stressed*. In English, lexical stress is specified in the lexicon. Among syllables which do not carry primary stress, speakers distinguish syllables which *can* be stressed, and those that are inherently *unstressed*, e.g. *Chi'nese/'Chi nese* versus *'ta ble/* ta'ble* (though this may not be entirely categorical, Fear, Cutler & Butterfield 1995). Lexically stressed syllables have the capacity to be pitch accented (see Ladd 1996, pp. 46-51). As well as possible local pitch movement, accented syllables are marked with greater duration, intensity and shallower spectral tilt than unaccented syllables; as well as full, rather than reduced, vowels. However, experimental evidence is conflicting on whether these last phonetic cues can also reliably distinguish stressed and unstressed syllables in *unaccented* positions (Huss (1978), Sluijter & van Heuven (1996), Campbell & Beckman (1997), see review in Terken & Hermes (2000, pp. 90-101)). This could be because, as we see in the next chapter, the relevant distinction is not [± stress] and [± accent]; rather duration, intensity and local

[1] In any work on the interface between two areas of linguistic research, one must be introduced before the other. However, the discussion below will necessarily make reference to prosodic concepts. If the reader is not familiar with them, they are directed to Chapter 3, where they will be fully introduced and explained.

pitch movement are all cues to multiple levels of relative prominence. This is suggested by Bell, Jurafsky, Fosler-Lussier, Girand, Gregory & Gildea's (2003) corpus-based study, which showed that unaccented stressed syllables have distinct phonetic features in some contexts, e.g. slow speech or when the word is discourse new; but not in others, e.g. fast speech. In other words, these phonetic cues (especially duration and spectral tilt) are correlates of degrees of prominence, not lexical stress per se.

Among lexical items, certain types of words are much more likely to be prominent than others. 'Function' words (determiners, prepositions and pronouns) are often unstressed at the sentence level.[2] In fact, if they are stressed, an extra meaning can be implied, e.g. *Kate Moss is THE supermodel of the noughties* (cf. Bell et al. 2003). Furthermore, different classes of content words, e.g. nouns, are more likely to be accented (phrasally stressed) than others, e.g. verbs (Ladd 1996, p. 187). This is exploited in pitch accent prediction systems, where text-based features including part-of-speech perform nearly as well as combinations of text and acoustic features (Hirschberg 1993, Conkie, Riccardi & Rose 1999, Pan, McKeown & Hirschberg 2002, Chen & Hasegawa-Johnson 2004). These differences could be attributed to the relative informativeness of different word classes. However, a recent study by German, Pierrehumbert & Kaufmann (2006) seems to show the bias against accents on prepositions holds even given the expected accent position due to information status. We explore this further in our corpus study in Chapter 6.

There is a strong correlation between syntactic and prosodic phrasing. Prosodic breaks can sound strange in the middle of syntactic phrases (see review in Shattuck-Hufnagel & Turk 1996). For example, in the middle of the subject NP in (2.3) (from Shattuck-Hufnagel & Turk 1996, p.197):[3]

(2.1) (George and Mary give blood)

(2.2) (George and Mary) (give blood)

(2.3) * (George) (and Mary give blood)

(2.4) (George) (and Mary) (give blood)

However, the two are not isomorphic: for any given syntactic parse there are multiple prosodic phrasings that are perfectly acceptable, e.g. (2.1), (2.2) and (2.4). Moreover, phrasing is affected by non-syntactic factors such as speech rate (Shattuck-Hufnagel & Turk 1996). On the other hand, there is much research showing prosodic cues can be used for syntactic

[2]One noted exception is particles in phrasal verbs, e.g. *carry on* (Hirschberg 2002, p. 34).

[3]In this, and all future examples in this thesis, round brackets indicate prosodic boundaries. Examples not enclosed in such parentheses are not marked for prosody.

disambiguation (see reviews in Cutler, Dahan & van Donselaar (1997), pp. 159-171 and Speer, Warren & Schafer (2003)). For example, Schafer (1995) showed that speakers interpret the PP *from Alabama* as attaching to the NP *her friend* in (2.5), whereas in (2.6) they interpret the PP as attaching to the whole VP *phoned her friend.*

(2.5) (Paula phoned) (her friend from Alabama)

(2.6) (Paula phoned her friend) (from Alabama)

(2.7) (Paula phoned her friend from Alabama)

In examples such as (2.7), however, which should be ambiguous, listeners still preferred the VP-attachment reading. It seems that while prosodic cues can aid syntactic parsing, speakers are very inconsistent about actually using them; perhaps depending on the speaker's awareness of the ambiguity and the information status of the elements involved (Speer et al. 2003). Further, evidence from prosodic breaks is liable to be overridden if it conflicts with the context (Cutler et al. 1997, p.169). The location of pitch accents may also be used to signal attachment ambiguities (Schafer, Carter, Clifton & Frazier 1996, Schafer, Carlson, Clifton & Frazier 2000, Hirschberg 2002). For example, Schafer et al. (1996) showed that in sentences like *The detective eyed the entrance of the house that shows clear signs of damage*, the relative clause is more likely to be perceived as attached to *entrance* if this is accented, or to *house* if this is.

As we shall see below, it has been argued that both prosodic phrasing and prominence are important to signalling information structure. Evidence of independent lexical and syntactic effects on these signals is therefore highly relevant. It could be argued that these effects are not independent because the purpose of syntax is to help derive information/semantic structure, so they are facets of the same thing. However, this view can be difficult to reconcile with some of the evidence above. We return to this point in our corpus analysis in Chapters 6 and 7.

2.2 Information Structure

Information structure describes how the information conveyed in a discourse is structured. It is usually framed in terms of creating a *common ground* of propositions relevant to the context that both speakers believe to be true (Stalnaker 1978). Therefore, as each utterance is said, the speaker also conveys its information structure, i.e. how they intend each entity, predication, etc. to reference, alter and/or update the existing discourse structure. Below

we set out the basic concepts claimed to mark information structure, in particular F(ocus)-marking and focus projection. We show how the various discourse phenomena related to focus have been given a unified explanation within Alternative Semantics theory (i.e. *kontrast* marking). At each point, we set out how focus has been claimed to be marked prosodically, showing that disagreements in the literature and problematic cases may actually stem from misunderstandings about the nature of prosodic prominence.

In the next section we show how focus marking interacts with the marking of givenness. We will see that there are in fact two distinct notions of givenness in the literature, relative givenness and discourse givenness. While relative givenness is more straight-forwardly related to focus, both are relevant to us as they have been claimed to have independent effects on prosodic realisation.

We will then show that "focus" in fact comprises two dimensions of information structure, the division between *kontrast* and background, and between theme and rheme. We show that the latter is closely related to prosodic phrasing. We will then move on to the contentious issue of whether kontrast within theme is marked by a categorically distinct prosodic marker to kontrast within rheme. The issue is often conflated with the marking of contrastiveness in general. We show that the perception of the semantic categories to be explained is coloured by assumptions about the prosodic distinctions involved. This issue will be discussed in Chapter 3, and form the focus of the experimental work in Chapter 4.

2.2.1 Focus

Information structure describes the salience and organisation of information in a discourse. The principal mechanism to control the salience of information is the marking of focus, which in turn regulates the organisation of information. Below we set out the basic concept behind focus, and F(ocus)-marking; then show how this relates to organisation, i.e. through focus projection. We will see that it is standardly held that the basic relationship with prosody is that F-marked elements are pitch accented. However, this turns out to be problematic in many cases. We will further see that most of the literature on focus attempts to give a unified explanation of various related discourse phenomena, including wh-focus, interpretation of focus-sensitive adverbs and given/new status; so we will organise our discussion around these phenomena. However, as we discuss in Chapter 5, it is still a largely unexplored empirical question whether they can in fact be subsumed under the concept of focus. Finally, we discuss the added implication of focus marking in many cases, i.e. contrastiveness, and see how this has been given a unified explanation in Alternative Semantics theory. We show, however, that this notion may not fully account for variation is the prosodic marking of focus.

2.2.1.1 F-marking and Wh-Focus

The clearest examples of focus marking (and those most often given in the literature) are question-answer pairs. The basic idea is that the part of a response that answers the question, i.e. relates to the wh-phrase, is the *focus* and carries a pitch accent; while the part that is contained in the question itself is the *background*, as in:[4]

(2.8) Q: What did Arun buy?

A: (Arun bought a PORSCHE)
background *focus*

Since at least Jackendoff (1972, ch. 6), this has been formalised in terms of F-marking on syntactic phrases (e.g. Rochemont 1986, Krifka 1991, Rooth 1992, Krifka 2006). Syntactic nodes which are not F-marked form the *presupposition* of the utterance, i.e. well-defined and under discussion in the context of the discourse; whereas F-marked nodes contain what is *asserted* by the utterance (in relation to the presupposition). So, (2.8) would be represented thus (notation from Jackendoff 1972, p. 247):

(2.9) [[Arun][bought [a PORSCHE]$_F$]]

(2.10) λx[Arun bought x] is *well-defined/under discussion*

(2.11) porsche ϵ λx [Arun bought x]

The distinction between *presupposition* and *assertion* is a crucial part of the 'meaning' of focus. We will see that it gives a unified explanation for the information structural interpretation of related discourse phenomena, including question-answering, association with focus-sensitive adverbs and the given/new status of referents. However, as we show, the assumptions both that the relationship is between focus and accenting per se; and between F-marking and syntactic nodes, prove problematic in many natural language examples.

2.2.1.2 Focus Projection

The first challenge for syntactic F-marking theories is to explain the pattern of accents in focussed phrases. So in (2.8), a focus on the object was unproblematically marked by an accent on *Porsche*. However, at least in transitive sentences, a single accent on *Porsche* can also apparently 'project' focus onto the VP or even the whole phrase in the following contexts, though focus cannot be projected from the subject in the same way:

[4] In this and all future examples, CAPS indicate accents.

(2.12) What did Arun do?
(Arun bought a PORSCHE)

(2.13) What happened?
(Arun bought a PORSCHE)

(2.14) What happened?
* (ARUN bought a Porsche)

Originally, the accent on *Porsche* was taken to be a 'default' nuclear accent whose position was determined by cyclical rules applied to syntactic structures, with all other accents being *contrastive* (Chomsky & Halle 1968). It is now generally thought that 'default' accents are in fact meaningful, and mark *broad* focus (scope over S or VP), or *all-new* utterances. Nonetheless, after Halliday (1968), standard focus projection theories assume syntactically determined distributions of accents given particular F-markings. Selkirk's (1995) much cited account claims the following rules determine how F-marking can be projected upwards from an accent, with the outermost F-marked phrase being the Foc(us) of the sentence (see also Selkirk 1984).

(2.15) An accented word is F-marked.

(2.16) F-marking of the *head* of a phrase licenses the F-marking of the phrase.

(2.17) F-marking of an *internal argument* of a head licenses the F-marking of the head.

Thus, the F-marking on (2.8), (2.12) and (2.13) is as in (2.18), (2.19) and (2.20) respectively (ignoring marking of the determiner for simplicity's sake). Since the subject *Arun* is neither the head of the phrase, nor an internal argument of the head; focus cannot project from the subject in (2.14).

(2.18) Arun bought [[a PORSCHE]$_F$]$_{FOC}$

(2.19) Arun [[bought]$_F$ [a PORSCHE]$_F$]$_{FOC}$

(2.20) [Arun [[bought]$_F$ [a PORSCHE]$_F$]$_F$]$_{FOC}$

Most of these accounts start with the assumption that accents F-mark words. This has always been rather problematic given the appearance of optional, and in some cases obligatory, accents both inside and outside focussed constituents. For instance, a weaker accent on *Arun* would sound perfectly acceptable to many speakers, at least in the VP- and S-focus versions. In fact, Gussenhoven (1999*b*) claims that (2.13) would not be acceptable unless

there was an accent on *Arun*. He disputes what he calls *the extended view* of focus projection in favour of his account: that focus can only project from an argument to its adjacent predicate, citing experimental work backing up this claim (Gussenhoven (1983), see also Ladd (1996, ch. 5)).[5] Certainly, an accent seems to be required when the subject phrase is long, even with object-focus (cf. Beckman (1996, pp. 52-4), Ladd (1996, ch. 5)):

(2.21) Q: What did Arun's mother-in-law think?

A: (Arun's MOTHER-in-law DISAPPROVED)

Selkirk allows that F-marking may be associated with nuclear accents, while pre-nuclear accents may appear because of phonetic constraints, e.g. rhythmical reasons or phrasal strengthening, i.e. accents marking beginnings of phrases. The difficulty with this is in knowing when pre-nuclear accents do represent F-marking (i.e. are meaningful) and when they don't; and indeed whether there are consistent phonetic differences between meaningful pre-nuclear accents and meaningful ones. Intuitively, it is hard to tell; compare the following response to (2.21), which implies a contrast on *mother-in-law* (see further in section 3.2.2):

(2.22) (Arun's MOTHER-in-law DISAPPROVED)
(but his FATHER-in-law LOVED it)

If the relationship is between nuclear accents and F-marking, then setting out precisely what the 'phonetic constraints' affecting pre-nuclear accent placement are is something the semanticist can afford to set aside. However, if the relationship is with accenting per se, but this relationship is partial, and these constraints are not well-defined, such a theory of focus projection loses much of its verifiability and ability to be generative.

Likewise, the assumption that focus projection is syntactic leads to some rather complicated explanations with certain syntactic structures. For example, it has been noted that certain intransitive sentences are most naturally said with main stress on the subject to signal broad focus, e.g. (the second from Ladd 1996, p. 188):

(2.23) (my CAR broke down)

(2.24) ?? (my CAR broke DOWN)

(2.25) (his MOTHER died)

(2.26) ?? (his MOTHER DIED)

[5] Byrd & Clifton (1995) and Welby (2003) show experimentally that focus can project to the predicate, but assume it cannot project from the object to the subject.

Selkirk is forced to account for this in terms of the F-marking of the traces of *car* and *mother* respectively (which move to the higher clause), licensing the F-marking of the verb in each case. This is potentially supported by the comparison between these unaccusative sentences and other intransitives such as the following (from Ladd 1996, p. 188):

(2.27) (my brothers are WRESTLING)

(2.28) (my BROTHERS are WRESTLING)

(2.29) (Jesus WEPT)

(2.30) (JESUS WEPT)

However, under this account, there is no explanation as to why an accent sounds strange on *down* and *died* in (2.24) and (2.26) respectively. At the least, there is an added implication with these readings, e.g. in (2.24) that the hearer knows the speaker has a car, or they are annoyed this should have happened today of all days; in (2.26), that the hearer had just been speaking about *his mother* as if she was alive. There is no such implication with the additional accents *before* the focus in (2.28) and (2.30).[6]

Further difficulties arise for syntactic accounts in cases such as the following, where there is apparently acceptable variation in the accentual marking of broad focus (from Bolinger 1972, p. 637):

(2.31) a. I can't go with you...
(I've got too many THINGS to do)

b. (... too many things to DO)

In this case, either pattern seems equally acceptable, even though the syntactic structure is exactly the same. This type of example has been argued to show, most famously by Bolinger (1972), that accent patterns are determined by the relative *informativity* of the elements in a clause, not by syntactic constraints at all. We discuss this in section 2.2.2 below, and show that certainly *relative givenness* within focussed phrases proves problematic for syntactic focus projection theories. However, here we note that this explanation is more plausible in some cases than others, even among the examples we have already seen. For instance, in the context of (2.12), one may argue that *bought* is predictable from *Porsche*, particularly if it was known by the speakers that *Arun* didn't have a *Porsche* before. On the other hand, in some cases, the unaccented and accented elements appear to be equally informative. In a

[6]This phenomenon can be seen clearly with the relative acceptability of an accent on the predicate in broad focus sentences in SVO versus SOV clauses in German (see Wagner (2005) and discussion in Ladd (1996, pp.187-193)). However, the relevant structures are not very common in English.

sentence coming out of the blue, *Arun* in (2.13) seems as informative as *Porsche*, yet a sole accent on *Arun* does not convey broad focus. In the case of (2.23), *broke down* is arguably as informative as *my car*. In an ordinary person's life there seem to be a similarly limited number of things that can break down (among inanimate objects at least), e.g. *boiler*, *washing machine*, *bus*; to the number of things that could have happened to one's car, e.g. *got serviced*, *was stolen*, *went fast*. In any case, our theory needs to explain why certain distributions of accents within phrases lead to marked focal interpretations (i.e. narrow focus), whereas others do not.

In the next chapter we discuss a number of recent proposals (especially Ladd (1996, ch. 6) and Truckenbrodt (1995)) suggesting many of the difficulties just discussed disappear if we take the association to be between focus and phrasal prominence within prosodic structure, and focus projection to be constrained by prosodic phrasing, not syntactic projection rules.

2.2.1.3 Alternative Semantics and Focus-Sensitive Operators

As well as affecting the interpretation of whole utterances, the position of the accent changes the interpretation of certain adverbs, such as *only*, *always* and *even*, with regard to their truth-conditional implications, i.e. the propositions in the *common ground* that are compatible with them.[7]

(2.32) A few months later, Arun and his friend Joel discussed their recent holidays...

(2.33) (Only Arun had DRIVEN to Paris)
Joel took the plane.

(2.34) (Only Arun had driven to PARIS)

a. Joel went to Leeds.

b. * Joel took the train.

An accent on *driven* implies the assertion is to do with the mode of transport, while an accent on *Paris* implies it is to do with the destination, so the rejoinder in (2.34b) sounds strange. This is consistent with the implications of F-marking on *driven* and *Paris* respectively.[8] However, there is an added meaning with such utterances, often described in terms

[7]Other expressions have been claimed to have similar effects, e.g. modals (*must*) and connectives (*if-then*) (see summary in Rooth 1996*a*). We restrict our discussion to adverbs, as the literature on them is more extensive and the prosodic issues much the same.

[8]Some theorists have argued that the association with focus-sensitive adverbs is resolved lexically or pragmatically (e.g. Partee 1999), as not all such adverbs behave the same way with respect to the truth-conditional effects on the common ground (Beaver & Clark 2002, Beaver & Clark 2003). As this debate does not involve prosodic effects, we leave it aside. A more serious challenge comes from non-pitch based second occurrence focus marking, which we address below.

of *exhaustivity* or *scalar implicature*. That is, in (2.34), the speaker picks out the focus *Paris* as opposed to a contextually determined set of alternatives, e.g. *Leeds*. In fact, these implicatures can arise from pitch accenting, especially with *emphatic* or *contrastive* accents, without the adverb being present. For example, the following alternative rendition to (2.34), with a particularly exaggerated accent on *Paris*, may also imply that *Arun* is luckier than *Joel*, as *Paris* is a better destination than *Leeds*.

(2.35) (Arun had driven to **PARIS**)

The distinction with an overt marker such as *only* is that the implicature seems to be cancellable; although it is disputed to what degree this is true if an *emphatic* or *contrastive* accent is used.

This type of implicature was actually originally argued to only result from *contrastive* accents, with all other accents being 'default', as was laid out in the last section (Chomsky & Halle 1968). However, as was pointed out by Bolinger (1961), any focus can theoretically be contrastive (Bolinger 1961, p. 87):

> Clearly in *"Let's have a PICNIC"*, coming as a suggestion out of the blue, there is no specific contrast with *dinner party*, but there is a contrast between picnicking and anything else the group might do. As the alternatives are narrowed down, we get closer to what we think of as a contrastive accent.

This insight has been formalised in the now widely accepted theory of Alternative Semantics (Rooth 1992). Rooth claims that the effect of F(ocus)-marking is to introduce, in addition to the ordinary semantic meaning of a proposition, a free variable which is the set of alternatives available to the focussed phrase in the proposition it appears in. Vallduví & Vilkuna (1998) uses the term *kontrast* to describe this definition of focus, which we adopt from now on.[9] So, taking example (2.8) again, the ordinary semantic value is given in (2.36), and the kontrast semantic value in (2.37):

(2.36) $[\![\ [_S$ Arun bought [a Porsche $]_F]\ ]\!]^o$ = { bought(Arun,porsche) }

(2.37) $[\![\ [_S$ Arun bought [a Porsche $]_F]\ ]\!]^f$ = { bought(Arun, x) | $x \in E$ }, where E is the domain of buyable things

Each focussed phrase α has a kontrast semantic value $[\![\alpha]\!]^f$ that introduces a free variable Γ which is restricted by the formula $\Gamma \in [\![\alpha]\!]^f$. Rooth convincingly analyses the various focus-related discourse phenomena in terms of the resolution of this free variable in the

[9]This is to distinguish the Alternative Semantics definition of focus from the more general notion of focus described above and the varying definitional and pragmatic uses of the word *contrast*. Note that we will continue to talk about F-marking, however, as this is the term almost universally used in the literature.

preceding context in a similar way to anaphora resolution (see discussion of given/new effects in next section).[10] For example, the question which prompted (2.8) has an ordinary semantic value thus:

(2.38) $[\![\ [_S\ [\ \text{What}\ ]_F\ \text{did Arun buy}\ ]\ ]\!]^o = \{\ \text{bought(Arun, } x) \mid x \in E \wedge \text{buyable}(x)\ \}$

Since the ordinary semantic value of the question phrase, which can be represented as $[\![\beta]\!]^o$, is an element of $[\![\alpha]\!]^f$, and it is available, it acts as an antecedent for the free variable, licensing the F-marking.

Similarly, he argues that adverbs like *only* introduce a universal quantification over properties. Taking example (2.34) again, the quantification obtained in the configuration (2.39) is (2.40). This says that if P is a property in a certain set of properties C, and if Arun has the property P, then P is identical to the property expressed by VP. That is

(2.39) $[_S$ Only Arun $\text{VP}_F\ ]$

(2.40) $\forall P\ [[\ P \in C\ \wedge\ P(\text{arun}) \rightarrow P = VP'_F\ ]]$

The role of kontrast is to identify the set C serving as a domain of quantification: the variable is set equal to the kontrast semantic value of VP. So the VP in (2.34) has a kontrast semantic value as in (2.41); which produces the interpretation of the utterances as in (2.42):

(2.41) $[\![\ [_{VP}\ \text{drove to}\ [\ \text{Paris}]_F\ ]\ ]\!]^f = \{\ \lambda x\ [\ \text{drove}(x,y)]\ |\ y \in E\ \}$

(2.42) $\forall P\ [[\ P \in C\ \wedge\ P(\text{arun})\ \rightarrow\ P = \lambda x\ [\ \text{drove}(x,\text{paris})\ ]]]$

This says that *Arun* has a property 'driving to x'. The value of this property is 'driving to Paris', among the contextually available set of appropriate alternatives of places he could be driving. That is, the alternative set is values of this property, not alternatives to *Arun* himself. Rooth claims that if we analyse *only* (and its attendant semantics (2.39) and (2.40)) as another discourse phrase having the property $[\![\beta]\!]^o \in [\![\alpha]\!]^f$, we can unify the analysis of question-answer congruence and *only*-association.

It is easy to see how an implicature of exhaustivity can arise given the concept of alternative semantics. If the set of alternatives itself is strictly limited from the context, exhaustivity follows naturally from the kontrast semantic value. Rooth also argues that the scalar reading,

[10]There is considerable debate within the semantics literature about the constraints on the resolution of this free variable; with some arguing that it is constrained by syntax (e.g. Krifka 1991, Krifka 2006), while others argue it is much more free (e.g. Rooth 1999, Büring 2004). As this does not seem to involve prosodic issues we assume the latter view. This accords more easily with the results of our kontrast annotation (see Chapter 5), showing antecedents can be very far away in a conversation or accomodated (see also link with the Rhetorical Structure Theory notion of contrast in Umbach 2004).

such as we saw in (2.35), arises through conversational implicature from the kontrast. The exaggerated accent on *Paris* implies that it is a better holiday destination than *Leeds*. Rooth represents this using a partially ordered set of alternatives (rather than the usual unordered set), where the higher member entails the lower:

(2.43) { go(Paris) > go(Leeds) }

This type of scalar implicature can act as one of the antecedents $[[\beta]]^o$ which licenses the free variable introduced by $[[\alpha]]^f$. We would therefore expect such an implicature to arise in situations where an appropriate overt antecedent is not available.

However, this brings us back to the difference between Chomsky's and Bolinger's explanations above, which we term the *availability* of the kontrast alternative set in the context. While Rooth's analysis nicely captures the idea that any focus can theoretically be contrastive, in some discourse contexts the make-up of the alternative set is very apparent, e.g. in (2.33) above the set is probably limited to {*drove*, *took the train* and *flew*}, and a kontrast on *drove* actively excludes the other options; whereas it seems unlikely the alternatives to *having a picnic* are actively available in Bolinger's scenario (or at least that the speaker's alternative set can in any way be said to be part of the common ground). Intuitively, the prominence of the focal accent is linked to the availability of the alternative set. So, a strong accent on *picnic* would be more consistent with the exclusion of *dinner parties* as the group activity. Equally, the scalar implicatures which Rooth claims arise in examples like (2.35) seem more available with strong or contrastive accents, i.e. on *Paris*. A 'neutral' reading does not necessarily imply this.

There are a number of proposals in the literature consistent with the idea that what we will call *restricted* kontrast is marked in a phonologically distinct way, either through pitch accent type or distribution. Firstly, there is the on-going debate about whether broad focus can be distinguished from object focus by the accent on the object, e.g. (2.8) versus (2.13) (we will return to this in section 3.2.4, but see Rump & Collier (1996)). Secondly, there are claims that accent type signals the implicatures discussed. Pierrehumbert & Hirschberg (1990) claim L+H* and L*+H invoke a scalar implicature, while H* does not. Ladd (1980) contains a similar proposal, linking his 'fall-rise' accent to the availability of alternatives, as opposed to 'fall' accents (see further section 2.3.1). Kiss (1998), in a comparison with Hungarian, and Gussenhoven (to appear), distinguish the marking of 'identificational' (our *restricted* kontrast) and 'informational' (*kontrast*) focus on the basis of pitch accent distribution (see also discussion in Umbach 2004). Others, including Rooth, maintain that the basic prosodic marking is of kontrast, *scalar* and *exhaustive* implicatures are not categorically marked, and are cancellable (unlike, e.g. the exhaustive implicature of *only*).

It seems to me that this debate is not primarily semantic but prosodic. There are at least three possibilities: firstly, there is a categorically distinct pitch accent type which acts unequivocally on top of F-marking implying *restricted* kontrast, similar to the effect of *only*. Secondly, the likelihood of a *restricted* kontrast reading increases with the prominence of the accent; or increased prominence could act in conjunction with pitch accent type. Lastly, these implicatures could result from general connotations of emphasis, or the types of illocutionary and affective 'meanings' discussed below, and not from focus per se. Put in this way, assessing the validity of the different theories above is difficult, but still an empirical question that can be tested on the basis of phonetic evidence.

2.2.2 Givenness

It is often observed that *new* entities tend to be accented, and *given* entities deaccented. For example, in (2.44), since *whisky* is mentioned in the first clause, it is deaccented in the second (from Ladd 1996, p. 175):

(2.44) I bought her a bottle of whisky, but it turns out...
(She doesn't LIKE whisky)

There are many exceptions to this, however, which are well-documented. These stem from both the definition of *given* and *new*, and from the association with accenting per se. For instance, the mention does not have to be explicit, in (2.45a), *the butcher* is *inferred* as referring to the person who did the operation; whereas in (2.45b), the referent is some unfairly maligned butcher (from Ladd 1996, p. 249):

(2.45) Q: Everything OK after your operation?
A: Don't talk to me about it! ...
a. (I'd like to STRANGLE the butcher)
b. (I'd like to strangle the BUTCHER)

Nor are all second mentions deaccented, in (2.46), accenting the second *movies* gives a marked interpretation, implying that the referents are different, i.e. not all movies qualify as movies (from Terken & Hirschberg 1994, p. 126):

(2.46) (There are MOVIES) (and there are MOVIES)

It is clear that a straight-forward definition of *given* as mentioned in the discourse, and *new* as novel is too strong. There are in fact two related, but distinct, notions of givenness in the literature. Most of the semantic work reviewed above defines givenness relative to the

current proposition (accounting for (2.46)). This definition stems from the relationship with focus-marking and, as we shall see, is most closely related to accent distribution. However, as we set out in section 2.2.2.2 below, there is a largely separate literature which defines degrees of givenness relative to the whole discourse (e.g. (2.45)). We will see that *discourse givenness* is also claimed to affect the prosodic realisation of referents.

2.2.2.1 Relative Givenness

We can see the idea behind relative givenness most clearly when looking at the interaction of focus marking in *question-answer* pairs and the givenness of referents. In general, there is a strong correlation between focal material and new material, i.e. the information being asserted in an utterance is likely to be *new*. However, answers are usually accented even if they are not *new* to the discourse; and the relative givenness of the elements in the answer phrase can affect accent placement:

(2.47) Arun looked round all the fancy car shops - Mercedes, Porsche, BMW, Lamborghini...
So what did he buy?
(Arun bought a PORSCHE)

(2.48) What colour did he get?
(Arun bought a RED Porsche)

(2.49) What did Joel buy?

a. (Joel bought a GREEN porsche)

b. * (Joel bought a green PORSCHE)

c. (Joel bought a green MERCEDES)

d. * (Joel bought a GREEN mercedes)

green Porsche and *green Mercedes* are both answers to (2.49). However, while accenting *Mercedes* in (2.49c) is compulsory (cf. (2.49d)), *Porsche* in (2.49a) needs to be de-accented (cf. (2.49b)).[11] This is not due to the givenness of *Porsche* in the discourse per se (as *Mercedes* is also given), rather its givenness in relation to the predication of *bought*.

In standard accounts, e.g. Selkirk's (1995), this is represented by F-marking at different levels of syntactic structures. So the F-marking of (2.49a) would be as in (2.50), and the F-marking of (2.49c) would be as in (2.51). The F-marking has to be independent of the

[11] Note that the felicity of (2.49b) and (2.49c) is not affected by the accenting of *green*. We assume this is accounted for by focus projection as discussed above.

givenness marking (at least in (2.50)), as the F-marking on *green* cannot project to the whole NP in the standard analysis.

(2.50) Joel bought [a [GREEN]$_F$ porsche]$_F$

(2.51) Joel bought [a green [MERCEDES]$_F$]$_F$

This is fine, except that F-marking has to be independently justified by question-answer congruence, and the relative givenness of the answer; which a unified theory of F-marking is supposed to avoid. It is this type of example that led Schwarzschild (1999) to claim that it is an *absence* of F-marking that invokes an interpretation of *Givenness*, not the *presence* of F-marking which invokes an interpretation of *focus* (see also Ladd 1996, ch. 5). He gives the following definition of Givenness (from Schwarzschild 1999, p. 151):

(2.52) Definition of GIVEN:
An utterance U counts as GIVEN iff it has a salient antecedent A and:

a. if U is type e, then A and U corefer.

b. otherwise: modulo $\exists$-type shifting, A entails the Existential-F-Closure of U

Intuitively, this says that the everything that is given is entailed by a salient antecedent in the context. Schwarzschild (1999) claims that using the two constraints in (2.53), an analysis of examples like (2.49) can be given that preserves the unified treatment of given/new deaccenting and question-answer congruence (in Optimality Theory terms).

(2.53) 1. **GIVENness** If a constituent is not F-marked, it must be GIVEN

2. **Avoid F** F-mark as little as possible, without violating GIVENness

Briefly, the Sentence:

(2.54) [*Joel bought a* [*GREEN*]$_F$ *Porsche*]

is Given because it is entailed by:

(2.55) $\exists x[bought(joel, x)]$

which is the question. The VP is given because:

(2.56) [*bought a* [*GREEN*]$_F$*porsche*]

is entailed by:

(2.57) $\exists x[bought\ a\ x\ porsche]$

However, *green* is not entailed by anything, therefore it is the only word that can be F-marked, and is accented (see the paper for more details).

Schwarzschild's analysis does effectively overcome some of the puzzles of accent distribution and focus projection. However, especially in regard to our discussion above about *restricted* kontrast, it does not seem intuitively appealing that our interpretation of utterances should arise solely from the lack of accents, rather than the accents themselves. We have seen that in some cases, *de-accenting* seems to have definite effects on the interpretation of an utterance, consistent with Schwarzschild's Giveness-based constraint, e.g. (2.45), (2.48), and (2.49); however, in other cases *accenting* seems to bring about a particular interpretation, consistent with Rooth (1992), e.g. (2.35) and (2.46); while other examples admit of either interpretation. That is, while marking a referent as kontrastive, and marking it as relatively given may theoretically be complementary (see analysis in Wagner 2006), there are subtle interpretative differences between the two. Though there is very little discussion on this in the literature (but see Ladd 1996, ch. 6), it seems to me the desire to unite the above phenomena in the first place stems from the assumption that *de-accenting* and *accenting* are complementary phonological processes. This is not necessarily so. As we develop in the next chapter, if we allow for multiple levels of prosodic prominence beyond [± accent], we can link degrees of prominence to the availability of the alternative set (as described above), something that is not usually considered in these accounts. In other words, one's theoretical description of the semantic facts is directly affected by one's understanding of the phonetic reality.

The same sorts of issues arise in the interaction of given/new status and the interpretation of focus-sensitive adverbs. So, in (2.58), *Porsche* is given, therefore de-accented, even though the restrictor of only is *an old broken-down Porsche*.

(2.58) I thought Joel had a Porsche too...
(Ah) (but he only has an OLD BROKEN-DOWN Porsche)

However, cases where the entire focus of such an adverb is given cannot be accounted for within focus to accent theories, i.e. the so-called *second occurrence focus* cases. As we can see in (2.59), there is a pitch accent on the focus associated with *even*, i.e. *mother*, but not on that associated with the second *only*, i.e. *card* (Beaver, Clark, Flemming, Jaeger & Wolters 2004, p. 46).

(2.59) Kate usually gets lots of nice presents on her birthday. But her brother only gave Kate a card today.
(Even her MOTHER only gave Kate a card today)

Proponents of pragmatically or lexically determined theories of *only*-association claim that since focus is marked by a pitch accent, and there is no (obligatory) accent on *card* in examples such as (2.59), then focus-marking cannot explain association with adverbs like *only*. However, recent work by Beaver et al. (2004), stemming from insights in Rooth (1996*b*), has showed experimentally that while such foci might not be marked with pitch accents, they do have higher intensity and duration than equivalent unfocussed material. They have argued that this marking is post-nuclear non-pitch-based prominence. Therefore, the focus is always marked, but is associated with the largest prosodic prominence in the scope of the adverb, not pitch accenting per se. We return to this claim in the next chapter.

Once more, general constraints on prosodic structure also prove problematic for the assumption that focus/relative givenness is marked by (de)accenting. Take the following example (adapted from Wagner 2006, p. 10):

(2.60) Last week after the game all that happened was that the coach praised John. I wonder what'll happen after this week's game...

a. (PROBABLY) (the coach'll praise JOHN)

b. * (PROBABLY) (the COACH'll praise John)

c. (PROBABLY) (the coach'll praise John AGAIN)

The whole clause *the coach'll praise John* is given in (2.60a) in relation to *probably*, having just been mentioned. This can be seen in the comparison with (2.60c). However, since every phrase has to have an accent, *John* is accented. Since the whole phrase is given, there is also no obvious reason why *coach* can't equally be accented (cf. (2.60b)).[12] We will see in the next chapter that these sorts of examples can be quite easily explained if focus is taken to be signalled by prosodic structure itself, not accenting.

2.2.2.2 Discourse Givenness

In most of the work just reported, the givenness of a referent in relation to the whole discourse is assumed to be essentially peripheral to the accenting, or prominence, patterns in an utterance (though a referent that is relatively given will usually also be discourse given). However, there is a largely separate body of research which relates the prosodic realisation of referents to their accessibility, informativity and/or predictability in the whole discourse. Given the gradient nature of these concepts, such accounts tend to view the relationship with prosodic marking rather differently. Rather than information status being directly stipulated

[12]Note that this example works given the 'default' accenting patterns in transitive sentences. As discussed in section 2.2.1.2 above, a similar result would obtain with regard to the 'default' accent on the subject in unaccusative sentences.

by accenting or prominence; the *probability* of a word being prominent (or having a particular accent type) is affected by some or all of these influences; or prosodic prominence varies *gradiently* in relation to information status. Some researchers even claim that focus is an epiphenomenon arising from these competing influences.

Established work on written language has shown the *accessibility* of an entity strongly influences its textual reference, e.g. pronoun use and definiteness (Grosz, Joshi & Weinstein 1983, Ariel 1990, Gundel, Hedberg & Zacharaski 1993, Grosz & Weinstein 1995). Accessibility broadly refers to how easy it is for a speaker to retrieve a referent in a text, measured by factors such as time since last mention, and the syntactic position of the last mention. In a game task, Terken & Hirschberg (1994) showed that subjects regularly de-accent previously mentioned nouns in the same syntactic position; but only sometimes de-accent nouns mentioned in a similar surface position, e.g. PP vs Object; and rarely de-accent in a different place and syntactic position, e.g. Subject vs Object. In an instruction-giving experiment, Watson & Arnold (2005) found that higher prominence rating, intensity, mean pitch and duration all lessened gradiently across the conditions: new, mentioned as location, mentioned, mentioned twice; though accenting was not affected (phrases were short and 90% of nouns were accented). Bard, Anderson, Sotillo, Aylett, Doherty-Sneddon & Newlands (2000) found a general relationship between intelligibility and accessibility. In a study of repetition in a map task corpus, they found that repetitions were generally less intelligible, whether or not the introduction had been said by the same speaker, or whether the hearer apparently understood the reference. They conclude that their results show the effect of givenness on reduction is explained by fast priming processes dependent on the speaker's knowledge, and only to a much lesser extent by inferential processes stemming from a speaker/hearer model (as is implied by relative givenness in the previous section). The study could not confirm the acoustic correlates of intelligibility, except to note that it is not straight-forwardly related to duration, or to accenting (Bard & Aylett 1999).

In recent work, Baumann & Grice (2006) have investigated the acceptability of different accentual markings given different types of *inferential* accessibility (cf. Prince 1981, Prince 1992). Specifically, they looked at *converseness*, e.g. *sister – brother*; *synonymy*, e.g. *lift – elevator*; *part-whole*, e.g. *hand – finger*; *hypernym-hyponym*, e.g. *flower – lily*; and *scenario*, e.g. *trial – judge* relationships, as well as whether the referent was *textually displaced* from its antecedent. In a perception experiment, subjects were asked to rate the acceptability of the accent on the referent, i.e. H*, H+L* (equivalent to English !H*) and no accent, given each of these conditions. They claim their results show a general scale between inactive and active accessibility corresponding with a scale of preference for H* over H+L* over no accent (see Table 2.1). However, as they note, the scale from H* to H+L* to no accent

Type of Accessibility	Pitch Accent Type Preferences	Deaccentuation of Target
converseness	no accent >> H+L* > H*	higher preference
part-whole	no accent >> H+L* >> H*	
synonymy	no accent >> H+L* > H*	
hyponym-hypernym	no accent >> H+L* >> H*	
hypernym-hyponym	no accent >> H+L* > H*	⇕
textually displaced	no accent = H+L* >> H*	
whole-part	H+L* >> H* = no accent	
scenario	H+L* > H* = no accent	lower preference

Table 2.1: Summary of results from Baumann & Grice (2006) ('>>': highly significant preference; '>': significant preference; '=': no significant preference).

Accent Type	Information Status
H*	New
L+H*	Addition of a new value
!H*, H+!H*	Accessible
L*+H	Modification of a given referent
L*, no accent	Given

Table 2.2: Information status and accentual marking, from Baumann (2005, p. 156).

is one of reducing peak height; and could also be interpreted as a reduction in prosodic prominence correlating with an increase in referent accessibility. This becomes even more clear from the discussion in Baumann (2005), where he relates these findings to the literature on information structure marking, proposing a general scale of referent marking for English and German, see Table 2.2. We will see possible examples of this type of effect in Chapter 7.

Bolinger (1972) suggests, in his well-cited piece, that it is relative semantic *informativeness*, rather than focus and syntactic projection, that determines the accenting pattern of words in an utterance, as discussed in section 2.2.1.2 above. So *crawling* in (2.61) is more informative than *things*, whereas *insects* is more informative than *crawling* in (2.62), and so on (from Bolinger 1972, p. 636):

(2.61) Those are CRAWLING things

(2.62) These are crawling INSECTS

(2.63) He was arrested because he KILLED a man

(2.64) He was arrested because he killed a POLICEMAN

Intuitively this idea has merit, although it does not always seem to hold. As we saw in section 2.2.1.2, there are many examples where different elements seem to be equally informative, yet there is a clear difference in acceptable accentual patterns. There are other examples where the normal accenting pattern clearly contradicts this explanation, e.g. in *I gave him five POUNDS* the default stress is on *pounds*, which is surely less informative than *five* (in the UK) (from Ladd 1996, ch. 5). Further, it is difficult to quantify *informativeness*, though some recent work has tried. Pan & McKeown (1999) reports significant improvement in pitch accent prediction, even using rather crude measures, such as the negative log likelihood of a word in a given corpus and the TF*IDF (term frequency-inverse document frequency, Salton 1989), i.e. the frequency of a word in a document relative to its frequency in all documents. They also report that "semantic abnormality", a subjective rating of a word's unexpectedness, is correlated with pitch accenting and phrase breaks (Pan et al. 2002). In a second game task experiment, Watson & Arnold (2005) tease apart the effects of predictability and informativeness, reporting higher informativity is correlated with higher mean $f0$ and prominence ratings, while predictability is correlated better with increased duration.

Finally, predictability is also found to be a good predictor of pitch accenting (Pan et al. 2002). There is further a strong correlation between predictability and the duration and care of articulation of words. Bell et al. (2003) and Bell, Brenier, Gregory, Jurafsky & Girand (2004) show word duration in the *Switchboard* corpus is consistently affected by predictability (bigram and unigram), controlling for a multitude of other factors including speech rate, position in phrase, disfluencies, pitch accenting and speaker characteristics. Further, speakers are more likely to use reduced forms in highly predictable contexts. This correspondence led Aylett & Turk (2004) to propose that the purpose of prosodic prominence is to control care of articulation/duration in order to smooth the redundancy in the speech signal as a whole (as per information theory, Shannon 1948). Both highly predictable and carefully articulated (and long) words are more likely to be recognised. Therefore, it makes sense to carefully articulate and lengthen unpredictable words and shorten predictable words. In a corpus of direction-giving monologues, Aylett & Turk (2004) show that $f0$ variation and language redundancy can account for about 60% in the variation in syllable duration (controlling for prosodic boundaries). The remaining 40% could in part be explained by the inexactitude of the measures of language redundancy used (log word frequency, syllable trigram and prior mentions), or suggest a further independent role for prosodic prominence. However, the results are broadly consistent with their theory.

Some of the studies reported above could be explained in terms of correlations between gradient measures of predictability and informativeness and semantic categories such as focus, and therefore could be accomodated within theories of the latter. However, other studies are well controlled and detailed enough to demand further consideration. We will return to the implications of this evidence in the discussion at the end of Chapter 6.

2.2.3 Theme/Rheme Structure and Contrast

So far we have seen that there are two motivations for focussing, the marking of the assertion (question-answering), and the marking of kontrast/relative givenness, the latter of which can operate inside the former. In the last section we discussed proposals that attempt to unify these foci, here we see cases where they need to be treated separately. In (2.65), *Moana* and *Geoff* are *kontrasted*, as defined above, in that each is in the alternative set of the other. Equally, *Paris* and *Brussels* are kontrastive, in the alternative set of places *Moana* and *Geoff* could be going by train. However, *Moana* and *Geoff* are also *pre-supposed*, i.e. they are 'topical', linking each clause to the preceding context.

(2.65) Moana and Geoff met at the train station...

(MOANA	was going to	PARIS)			
Kontrast	*Background*	*Kontrast*			
(*theme*)	(	*rheme*	)		
(and	GEOFF	was going to	BRUSSELS)		
	Kontrast	*Background*	*Kontrast*		
	(*theme*)	(	*rheme*	)	

Below we see that this type of example can best be accounted for by defining the effect of *focus* in two dimensions, i.e. the kontrast/background division we have seen, and the distinction between theme and rheme units. We then move on to the contentious issue of whether, and how, kontrast within the theme is distinguished from kontrast within rheme. We will see that this is closely related to the debate set out above as to whether there are distinct 'contrastive' accents.

2.2.3.1 Theme/Rheme Structure

In examples like (2.65), we can clearly see the effect of one type of 'focus-marking' working inside the other. This type of evidence has led to the conception of information structure on two dimensions. One dimension, called the *informational domain* by Vallduví & Vilkuna (1998) encodes the status of an entity or property in relation to the current discourse model, i.e. whether the element relates back to what has already been said, the *theme*

(*Moana/Geoff*), or serves to advance the discourse, the *rheme* (*go to Paris/go to Brussels*).[13] There are many varied, and overlapping definitions of theme and rheme in the literature. However, as shown by Kruijff-Korbayová & Steedman (2003), many of these come down to the level at which one believes the distinction applies, i.e. to the utterance, sentence, clause, prosodic phrase or recursively at any level of discourse structure. We will see in the next chapter that this definitional dispute can be straight-forwardly resolved by a more flexible conception of prosodic phrasing than most works assume. The second dimension, called the *quantificational* domain by Vallduví & Vilkuna (1998), relates to how salient an element is in the discourse, i.e. the distinction between *kontrast* and *background* discussed above.[14] Kruijff-Korbayová & Steedman (2003) show that the theme/rheme dimension is closely related to the linguistic structure of Grosz & Sidner's (1986) model of discourse structure; while the background/kontrast dimension is akin to their attentional structure, showing the applicability of these concepts to a computational model of discourse.[15]

The *theme/rheme* division strongly constrains prosodic phrasing. Steedman (2000) claims all prosodic boundaries occur at information structural boundaries, though not all such boundaries are prosodically marked (see also Truckenbrodt 1999, Büring submitted). For example, (2.66) shows his postulated realisation of the first clause in (2.65) using ToBI notation (see description in section 3.1.3). Note that an LH% boundary after *going* would be equally acceptable. It is often difficult to define the boundary between *backgrounded theme* and *backgrounded rheme*. However, as Steedman argues, this probably stems from indeterminacy in the semantics, i.e. whether the speaker takes *going somewhere* to be entailed by being in a train station.

(2.66) (MOANA) (was going to PARIS)
L+H* LH% H* LL%

Steedman (2000) argues that this theme/rheme division actually constrains syntactic parsing, within his Combinatory Categorial Grammar (CCG). That is, the order in which constituents of a sentence are combined in deriving the Logical Form is constrained by prosodic phrasing. So, in the above example, *was going to* would be combined with *Paris* first, but if it was phrased with *Moana*, these would be combined first (see paper for further details). As we show in the next chapter, Steedman's claim probably needs to be qualified with regard to independent effects on prosodic phrasing, such as speech rate (e.g. the *George and*

[13]This division is also called topic/comment and topic/focus (see Kruijff-Korbayová & Steedman 2003).

[14]Also called background/focus, presupposition/focus, context bound/unbound (see Kruijff-Korbayová & Steedman 2003).

[15]Grosz & Sidner (1986) outline a third dimension of discourse structure, *intentional structure*, which is analogous to the dialogue acts described in the next section.

Mary give blood phrasings in section 2.1); although we will see that theme/rheme structure is marked by prosodic phrasing at some level of phrasing structure.

This insight helps resolve some of the difficulties of traditional syntactic focus projection accounts in dealing with apparently independent motivations for F-marking (e.g. the *green Porsche* sentences in (2.49)). That is, the problem of determining the scope of focus (Selkirk's (1995) Foc-marking) is in fact the problem of determining the scope of theme/rheme units. If we look at the *Arun/Porsche* sentences in (2.8), (2.12) and (2.13) which motivated the need for focus projection in the first place, the scope of the focus is directly determined by the presupposition contained in the question, i.e. it is the *rheme*. So the claim here is that theme/rheme boundaries are determined by prosodic phrasing, which in turn constrains syntax, rather than being directly constrained by syntax. This, of course, raises the problem of defining the scope of the kontrast within the theme or rheme phrase, i.e. which parts of the theme or rheme form part of each alternative set. Steedman assumes that kontrast is directly marked by pitch accents within these theme/rheme units. However, as we saw in the last section, the association with pitch accenting per se is problematic (see e.g. the acceptability of 'other' accents before and after the focus in (2.24), (2.28) and (2.49); the marking of given foci in (2.59); and interaction with prosodic constraints in (2.60)). Further, we suggest in the next chapter that, in longer theme/rheme phrases, degrees of prominence may be linked to the availability of different properties in the alternative set. For example, in (2.21) above, both *Arun* and his *mother-in-law* could potentially be kontrastive, i.e. as opposed to *other people's mothers-in-law* and *other relations of Arun* respectively. We suggest the interpretation depends not only on contextual appropriateness, but the prominence of both. We will see that this analysis neatly contains cases of so-called 'nested foci' within a two dimensional information structure (cf. Féry & Samek-Lodovici 2006).

2.2.3.2 Thematic Marking and Contrast

As we saw in section 2.2.2.1, themes are usually given and therefore unaccented (cf. Gundel 1985), so the theme forms a single phrase with the rheme. However, as we just showed, when there is a kontrast within both the theme and the rheme, they tend to form separate phrases. Since at least Bolinger (1965, pp. 57-66), it has been claimed that these themes are marked with a different pitch accent type than rhemes. Theme accents are often claimed to sound 'scooped'. Jackendoff (1972) claimed B-accents, i.e. 'rise-fall-rise', mark topics (our themes), while A-accents, i.e. 'fall', mark foci (our rhemes).[16] Steedman (2000) identifies this as the distinction between L+H* (LH%) and H* (LL%) respectively (see (2.66)), while

[16]Note that while Jackendoff adopted Bolinger's terminology, his phonetic descriptions of A and B accents are different. In Bolinger's scheme, the B accent is more akin to a 'secondary', or non-nuclear, accent.

	Theme	Rheme
No implicature	Backgrounded	Plain Kontrast
Restricted Alt Set	Kontrastive Theme	Contrastive Rheme

Figure 2.1: Two conceptions of the marking of 'contrast' assumed in the literature: *contrastive accents* (red) versus *theme/rheme kontrast* (dashed blue). Note that the 'Backgrounded' category does not take part in the accent type division indicated by the lines.

thematic kontrast is claimed to be marked by 'fall-rise' in Büring (2003), and (L+)H* LH% in Büring (submitted) and Oshima (2002).

This phonetic description is very similar to that of 'contrastive' accents in section 2.2.1.3 (see Ladd 1980, Pierrehumbert & Hirschberg 1990), which we claimed to mark *restricted* kontrast. Indeed, in many accounts, the marking of contrastiveness and thematic kontrast is collapsed (e.g. Krahmer & Swerts 2001, Watson, Tanenhaus & Gunlogson 2004). While thematic kontrasts tend to also be contrastive (since themes are normally given), it is possible to separate these notions. 'Contrastive' interpretations are possible within the rheme too, e.g. in (2.65). In most accounts a single distinction between either theme and rheme kontrast, or between contrastive and not-contrastive, is assumed, see Figure 2.1.[17] However, as we will see in later chapters, particularly Chapter 4, if they are separated, they seem to have distinct prosodic effects. It appears that the number of categories that one assumes to be separate in this table stems from the number of categories one believes to be prosodically distinct.

One further difficulty is that, as will be discussed at length in the next two chapters, experimental evidence both for a categorical distinction between L+H* and H* (see Ladd & Schepman 2003) and for a distinction between *contrastive* and *ordinary* accents in general is

[17]Note that in Steedman's analysis, to be described in the next section, quite a lot of weight is put on the analysis of certain whole propositions being thematic, i.e. so-called *isolated themes*. It follows from this that there could be a division between plain kontrast and contrastive readings within the theme as well. We have not included this here, as we believe the existence of such a division rests on whether one accepts his arguments about the implicatures arising from these isolated themes (and therefore their existence), and so is fairly internal to his analysis. Even this three-way division is not usually captured in the literature. Further, his theory would still require a categorical accent type distinction between themes and rheme in contrastive contexts, which we test directly in Chapter 4.

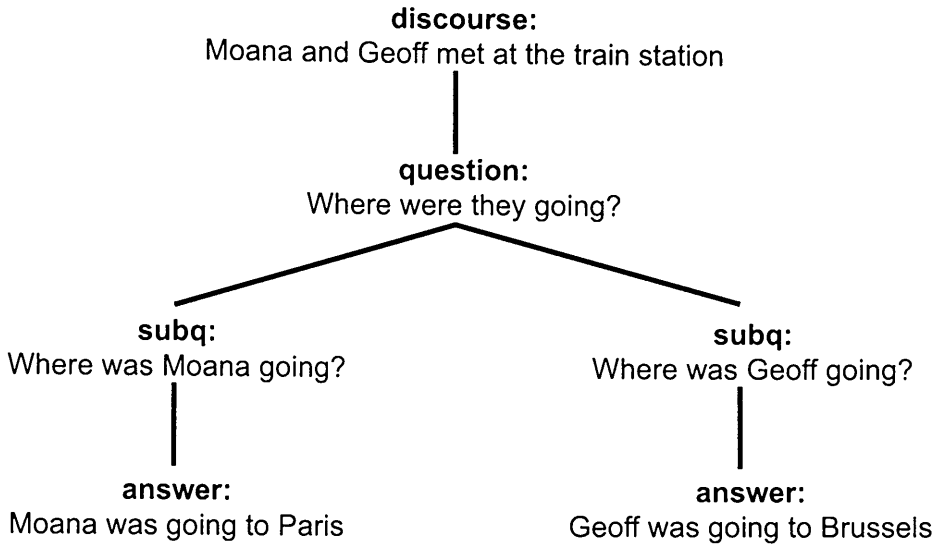


Figure 2.2: Question-Under-Discussion analysis of (2.65), (adapted from Büring 2003).

mixed. While most of the theoretical work (including Steedman) is essentially agnostic about the nature of the categorial distinction, the theory demands that it does reside somewhere. In Büring's (submitted) account, it is unclear if it is the pitch accent, boundary or overall tune that signals the distinction. However, if it is the boundary tone, then all the other illocutionary and affective 'meanings' of rising boundaries need to be accounted for (see sections 2.3.2 and 2.3.3).

Another unresolved issue is the semantic status of the distinction between theme and rheme. Steedman argues, drawing from his assumptions about the prosodic marking, that theme/rheme and kontrast marking form part of the surface structure representation, and are therefore used in the derivation of Logical Form. On the other hand, Büring (2003) sets out a *pragmatic* theory of the relationship (see also Büring submitted).[18] Drawing on Roberts (1998), Büring (2003) proposes that kontrastive themes (which he calls *contrastive topics*) indicate *strategies* for *moves* in a hierarchical discourse structure, showing how the speaker will navigate the sub-questions of the main *Question Under Discussion* (QUD), as in Figure 2.2. Kontrast within the theme works in the same way as Rooth's F-marking, i.e. it implies a set of alternative sub-questions of the QUD. Notions of *relevance* and *informativity* then ensure only appropriate accent marking for the context is allowed. As Büring points

[18]This is also closer to Jackendoff's (1972, pp. 262-3) original proposal.

out, his system is in practical terms very similar to Steedman's, the difference being whether kontrastive theme marking is formally part of the surface structure of the utterance. That is, his description is more akin to a instruction to the processor on the order to consolidate alternative sets in an utterance (see also McNally 1998).

2.3 Intonational Meaning

It is intuitively apparent that intonational tune is used to convey higher level 'meanings' about an utterance, e.g. its illocutionary status (*question*, *statement*), or affective connotations (*uncertain*, *polite*). Despite a wide body of research on the subject, it is still altogether less clear how these meanings arise from the intonational tune. As we will see below, stemming from the work just discussed have been a number of proposals claiming pitch accent and boundary tone type have a much wider role in signalling richer aspects of information structure. They further claim that illocutionary and affective connotations arise through conversational implicature from the compositional meaning of these tonal events. The difficulty with these proposals is in verifying that these meanings are in fact generalisable beyond the examples used by the researchers involved. Against this view is evidence that it is not individual tonal events, but whole tunes that convey the relevant meanings. We will see in the next chapter that such evidence could potentially be accommodated within the 'implicature from tones' approach. More problematic is evidence that these meanings arise from more gradient variation in the broader phonetic correlates of prosody. The pertinence of these competing explanations impacts directly on the study of information structure. Under the latter view, the relevance of illocutionary and affective connotations is as a further variable influencing the realisation of the prosodic signals of information structure; and consequently downplays the relevance of the tonal contour to our inquiry. Under the former, showing how illocutionary and affective connotations arise becomes an integral part of explaining and justifying tonal prosodic categories that signal information structure. In other words, one's view on this debate impacts upon how persuasive the claim is that basic information structure (theme/rheme) is signalled by tonal events.

Below we briefly set out proposals which claim a much greater role for pitch accent and boundary tone type in signalling richer aspects of information structure. We then go on to assess how well these proposals hold up given other evidence about how the illocutionary and affective implicatures they are meant to convey are signalled. We will see that the validity of this work depends in large part on one's belief in the categorial versus gradient nature of the phonetic variation involved. We suggest it also depends on the extent to which one believes information structure is conveyed by intonational tune.

	[C]	¬[C]
θ	L+H*	L*+H
ρ	H*, (H*+L)	L*, (H+L*)

Table 2.3: Meanings of ToBI pitch accents, in terms of theme (θ) and rheme (ρ) status and polarity in relation to the common ground (C) (re Steedman 2006*b*, p. 17).

[S]	L, LL%, HL%
[H]	H, HH%, LH%

Table 2.4: Meanings of ToBI boundary tones, in terms of the marking of speaker (S) versus hearer (H) supposition (re Steedman 2006*b*, p. 17).

2.3.1 Richer Information Structure

There have been a number of proposals arguing that much richer information about the propositional content of utterances is signalled by pitch accents and boundary tones. They are based around the idea that illocutionary and affective 'meanings' arise through conversational implicature from a number of basic properties of the information structure signalled prosodically.

Steedman has developed one of the most thorough and formally explicit systems (Steedman 2000, Steedman 2006*a*, Steedman 2006*b*). As discussed above, he adopts the idea that pitch accents mark kontrast; and further claims information structure strongly constraints prosodic phrasing, as well as that certain accents distinguish theme (θ) phrases (L*+H as well as L+H*) from rheme (ρ) phrases (L* as well as H*). But Steedman (2006*b*) goes on to claim specific meanings for boundary tones: falling boundaries mark elements the speaker regards to be his or her own supposition ([S]); while rising boundaries mark elements that the speaker regards to be the hearer's supposition ([H]), see Table 2.4. Finally, Steedman claims that the polarity of these suppositions in relation to the common ground ([C]) is also marked by pitch accent type. Using L+H* and H*, the speaker marks that unit should be added to the common ground, using L*+H and L* they mark that it should not (cf. Stone 1998, Stone 2004). The interaction with theme/rheme status can be seen in Table 2.3. Many of Steedman's examples have much intuitive appeal, take the following (from Steedman 2006*b*, pp. 19-20) (see the paper for more details).

(2.67) H: Is it raining?

S: I don't KNOW
H* LL%

$[S]\rho[C]\neg(know'raining'me')$

"I make it common ground that I don't know if it's raining"

(implies *I don't know if it's raining*)

(2.68) H: Is it raining?

S: I don't KNOW
L* LL%

$\neg[S]\rho[C]\neg(know'raining'me')$

"I do not make it common ground that I don't know if it's raining"

(implies *You should know I don't know, I don't care that I don't know*, etc.)

(2.69) H: Is it raining?

S: I don't KNOW
H* LH%

$[H]\rho[C]\neg(know'raining'me')$

"You make it common ground that I don't know if it's raining"

(implies *You know I don't know, Why ask me?*, etc.)

(2.70) H: Is it raining?

S: I don't KNOW
L+H* LL%

$[S]\theta[C]\neg(know'raining'me')$

"I supposed it to be common ground that I don't know if it's raining"

(implies *You already know I don't know*)

One of the most well-known proposals is Pierrehumbert & Hirschberg's (1990), which draws on many of the same intuitions as Steedman. H* marks items as to be added to the hearer's mutual belief space, usually as *new*. L* marks an item as salient, but not part of what the speaker is predicating. L*+H marks lack of speaker commitment to a proposed scale. L+H*, as was discussed in section 2.2.1.3, also evokes a salient scale, but to be added to the mutual belief space. Distinct meanings are claimed for phrase accents and boundary tones. L- separates the current phrase from the following one, while H- indicates they form a composite unit. Boundary tones, on the other hand, have scope over the entire phrase, and indicate whether it is 'forward-looking' (H%) or not (L%).

One last, older, proposal describes intonational meanings in terms of accent categories plus modifications of those categories.[19] Gussenhoven (1984) proposes that there are three basic nuclear tones in English, with functions as follows (from Gussenhoven 1984, p. 201-2):

1. **Fall** = *Addition:* Speaker may add the variable to the background
2. **Fall-Rise** = *Selection:* Speaker may select a Variable from the background
3. **Rise** = *Relevance:* Speaker may choose not to commit himself as to whether the Variable belongs in the background, used for relevance testing

A *fall* roughly corresponds to H* (or H*+L/H* L-), and its description is similar to Steedman's rheme. *Fall-rise* equates to (L+)H* LH%, which Büring (2003) claims to mark themes, and has a similar function here. The *rise* accent could be L*+H, which concurs with Steedman's claim that this accent marks elements that are not to be added to the common ground. So far, then, the proposal seems to arise from similar intuitions to those above. Gussenhoven goes on to claim that each accent type is subject to modifications, *delay*, *stylisation*, *half-completion* and *range*, which all add particular meanings to the basic accents (see paper for more details). *delay* and *range* turn out to be relevant to our argument in later chapters:

1. **Delay** = *Non-Routine, Significant:* Delaying the movements (or H and L location) associated with each accent adding an implication of non-routineness, or high significance, e.g. compare *I most CERTAINLY believe this is true* with H*, L+H* and L*+H respectively on *certainly*.
2. **Range** = *Insistence:* The insistence of the speaker on the given meaning increases as the speaker's range gets larger, e.g. compare *I said come HERE* to *I said come* HERE.

These modifications are not easily described in ToBI. *Delay* could correspond to a shift from H* to L+H* to L*+H, however, the parallelism of the effect would be lost. The other modifications are not captured at all.

As we will see below, although these proposals have intuitive appeal, is it difficult to test empirically in which direction this implicature can be said to go, i.e. are the basic properties fundamental or are they really meta-linguistic attempts to draw generalisations that arise from similar but distinct prosodic signals of such types of meaning? Further, there is the problem that some distinctions are tied to empirically disputed prosodic categories

[19] Ladd (1983) proposed a similar scheme around the same time, however his did not make such explicit claims about the meanings of each category and modification. The category meanings are also similar to Brazil's (1975) proposal. In particular, Gussenhoven's 'selection' is akin to his 'referring', and Gussenhoven's 'addition' to his 'proclaiming' (see also Brazil 1978, Brazil 1985).

(e.g. L+H* and H*), as we have noted. Gussenhoven's account serves to remind us that the concentration on the alignment of tonal targets in intonation description (particularly in ToBI) may have biased intuitions about prosodic meaning to be couched in these terms, missing more subtle manipulations of phonetic cues to prosody. We will see this in evidence set out below and return to this point in the next chapter.

2.3.2 Illocutionary Force

Intuitively, the intonational tune is important to signalling the illocutionary force of the propositional content of an utterance, i.e. whether it is a *declarative question*, *continuation*, *contradiction*, *request* or *statement*, etc. (Clark 1996), also called the *speech act* (Levelt 1989), or *dialogue act*.

For the purposes of exposition, we will look at the prosodic signals to the distinction between *declarative questions* and *statements*. Such questions are often observed to have lower pitch on the focus, and rising boundaries, so the statement in (2.71) might be distinguished from the question in (2.72) thus:

(2.71) We're going OUT tonight
H* LL%

(2.72) We're going OUT tonight
L* LH%

Pierrehumbert & Hirschberg (1990) claim that the question implicature in (2.72) arises from the combined meaning of the tonal events. The L* accent on *out* conveys that it is salient but does not need to be added to the speakers' mutual beliefs, i.e. because it is being questioned. L- marks the utterance as complete, turning over to the other speaker. H% marks it as 'forward-looking', i.e. the hearer needs to answer it. Steedman's (2006*a*) analysis would arrive at a similar result through a broader implicature. The gloss of this tune would be something like "you do not make it common ground that we're going out tonight". This gives rise to an implicature that the hearer needs to answer whether this proposition was in fact part of the common ground (and therefore didn't need to be asserted).

Earlier work suggested that such implicatures arose from the contour as a whole, rather than the meaning of each of the composite tones (whether or not the tune itself was believed to be composed of sequences of tones) (e.g. Pike 1945). For example, Sag & Liberman (1975) claimed that the '*tilde*-contour', as a whole, has the function of forcing an utterance to be interpreted as a question:

(2.73) 
We're going OUT tonight?

(2.74) 
* We're going OUT tonight?

This view was criticised at the time as being too 'brittle', i.e. ungeneralisable, for example Cutler (1977, p. 106) disclaimed "the attempt to extract from [intonation contours] an element of commonality valid in all contexts must be reckoned a futile endeavour". It is also hard to account for apparent similarities in meaning between similar tunes (e.g. see Pierrehumbert & Hirschberg's (1990) analysis of the *contradiction contour* and discussion in Ladd (1996, ch. 6)); and for independent constraints on where pitch accents are placed within an utterance, e.g. focal structure. These both fall out easily in the compositional approach. However, more recent evidence has shown that dialogue act type is signalled (at least in part) by the overall height and direction of $f0$ movement. In her study of declarative questions, Gunlogson (2003, p. 10) found that question intonation only needs to be "non-falling from the nuclear pitch accent to the terminus and ending at a point higher than the level of the nuclear accent", i.e. compatible with H* HH%, L* HH%, L* HL% and L* LH%, the generalisation between these being hard to capture in ToBI. This was confirmed by Šafářová & Swerts's (2004) experiment on the perception of declarative sentences from a corpus of spontaneous speech. Further, Eady & Cooper (1986) found that $f0$ declination does not seem to happen with questions, even following a focus (often said to be realised as H*L-), consistent with Gunlogson's (2003) analysis.

As we will discuss in section 3.3.2, these two positions are not necessarily incompatible if Pierrehumbert & Hirschberg's (1990) and Steedman's (2006*a*) assumption of the strict compositionality of tonal meaning is relaxed. That is, if intonational meaning is said to derive partly from the meaning of individual tones, and partly from semi-lexicalised intonational tunes. However, it does call into question the claim that these tones are simultaneously being used to convey basic properties of the organisation and salience of information in an utterance. That would place a very high informational 'load' on each tone.

Blithely unaccountable to such debates, work on the automatic recognition of dialogue acts has been reasonably successful using a variety of acoustic cues over whole phrases (Wright & Taylor 1997, Shriberg, Taylor, Bates, Stolcke, Ries, Jurafsky, Coccaro, Martin, Meteer & Ess-Dykema 1998). Results from Shriberg et al.'s (1998) decision tree classifier of dialogue acts in the Switchboard corpus (see further in Chapter 5) are revealing. They looked at the number of times each type of feature, i.e. $f0$ (including the overall contour),

intensity, duration, pausing and speech rate, was used in each tree as a measure of its usefulness. Overall, duration features were used in over 50% of decisions; $f0$, pausing and energy features were used about 10% of the time each. Interestingly, models built removing each type of acoustic feature successively (e.g. without duration or $f0$ measures) did not perform significantly worse than the full model, suggesting there is a lot of redundancy in the acoustic signal. In separating questions from statements, the most reliable cues were that statements were longer and had more pauses than questions, with the single predictor of change in speaking rate (speakers vary more in questions than in statements) out-performing $f0$ features. $f0$ features did work as expected, however, questions had higher mean $f0$ than statements and rising boundaries, while statements did not. Of course, a statistical decision-tree classifier is not a human being. However, such evidence may indicate that in spontaneous speech we attend to things like overall phrase length, pausing and speech rate variation much more than current theories of dialogue act signalling account for. This may not be captured in phonetic experiments which control for such factors.

Evidently, the relevance of the above work to our concerns depends a lot on how one believes information structure is signalled prosodically. If illocutionary force arises directly from conversational implicature inferred from the meaning of pitch accents and boundary tones, then it is inherently part of the study of information structure. At the other end of the spectrum if pitch accents (and relative prominence) are used to signal information structure, and illocutionary force is signalled by intonational tune and possibly other phonetic cues on entire phrases; then these 'meanings' do not concern us directly. Of course, there are many possible explanations lying between these two positions. The examples given by researchers such as Pierrehumbert & Hirschberg (1990) and Steedman (2006*a*) do have intuitive appeal. However, there have been surprisingly few studies testing the general applicability of their proposed 'meanings' outside the examples given in the relevant papers. In fact, it would be difficult to do so. The meanings of the different tonal events can seem so abstract that it is difficult to absolutely rule out or rule in the use of a particular tune in any given discourse context.

2.3.3 Affective Connotations and Emotion

There is no doubt that affective connotations ('conscious' attitudes a speaker shows towards what they're saying, e.g. *polite*, *uncertain* or *sarcastic*) and emotion ('involuntary' mental states, e.g. *happy*, *sad* or *angry*) are 'meanings' primarily conveyed by prosody.[20] Such

[20]The distinction between active *affective* and passive *emotive* mental states is often made in the literature, and is relevant here (see Scherer & Banziger 2004). However, it is rather orthogonal to our purposes exactly where the boundary, if there is one, between the two lies.

connotations (particularly emotive states) are often argued to be (semi-)universal, and signalled by *gradient* variation in prosodic cues, primarily pitch; and are therefore interpreted parallel to the linguistic signal (see Ladd 1996, ch. 1). However, it is also true that affective 'meanings' can be very fine-grained and subtle; to a degree that our current understanding of the manipulation of 'global' prosodic parameters does not really capture. It has therefore been argued, as we saw above, that such connotations arise through implicature from the information structural properties of pitch accents and boundary tones. Apart from the general indeterminacy of these claims noted above, such theories need to contend with evidence that these sorts of implicature arise from the *interaction* of different properties of the intonation contour and the linguistic signal, rather than from the tones themselves. Further, it is argued that global variation of pitch in itself interacts with the interpretation of the contour, i.e. because of the 'universal' correlates of high and low pitch.

For instance, in a series of studies, Ward & Hirschberg (1986) showed that the L*+H LH% contour is associated with *uncertainty* and *incredulity* (see also Ward & Hirschberg 1985). In Pierrehumbert & Hirschberg (1990), this is analysed as being an implicature created by the meaning of the L*+H accent: to convey lack of predication and evoke a scale.

(2.75) A: Alan's such a klutz

B: He's a good badminton player
L*+H L H%

So in (2.75), the speaker conveys they are uncertain whether this is counter-evidence to the proposition *Alan is a klutz* (from Pierrehumbert & Hirschberg 1990, p. 295). The speaker offers a piece of new information without predicating it, thus creating the uncertainty, and means this to be evaluated on the scale of *properties of klutzes*.[21] In order to uphold this analysis, Pierrehumbert & Hirschberg (1990) must claim the implicature holds even apart from the boundary tone, LH%. To my knowledge this has not been tested empircally; and their own examples are far from convincing. They claim the following example also conveys uncertainty (from Pierrehumbert & Hirschberg 1990, p. 295):

(2.76) A: We don't have any native speakers of German here. So let's work on Chinese.

B: Jürgen's from Germany
L*+H H H%

To me, this feels more like a contradiction or polite suggestion. There is no uncertainty as to whether *Jürgen is from Germany*, nor, unlike (2.75) above, whether this is relevant

[21] Pierrehumbert & Hirschberg (1990) are not explicit about what they mean by 'lack of predication'. However, for the purposes of the argument we will take this to mean that the proposition *he's a good badminton player* is not predicated in relation to the proposition *Alan's a klutz*, i.e. it does not explicitly negate A's assertion; rather than the predication of *a good badminton player* itself.

evidence. The full tune could convey uncertainty about whether A knows *Jürgen is from Germany*, although this does not fall out as easily. In the following alternative reply from B, it is hard to see how the alleged meanings (of uncertainty, lack of predication or a scale) is implied at all:

(2.77) B: I'd be happy to
L*+H L L%

Steedman's (2006*b*) analysis is more subtle. His gloss of the reply in (2.76) would be "you do not suppose it to be common ground that the theme is Jürgen is from Germany", which leads to the implicature that A should take this fact on as part of the common ground, but is more polite than if B had asserted it. So his claim would be that this reply does not give rise to an implication of uncertainty, but of a polite contradiction, which is appropriate in the context. His gloss of (2.77) would be "I do not suppose it to be common ground that the theme is I'd be happy to work on Chinese". This leads to the implicature that A should accommodate this in the common ground, with the added implication that this fact may have been in dispute.

As we said above, in some cases Steedman's analysis can seem reasonably persuasive. However, his analysis of these sorts of utterances as "isolated themes" does greatly expand accepted ideas about the notion of thematicity (as is noted by Steedman 2006*b*, p. 35). Further, we would argue, the main evidence for this expansion is intuitive ideas about the implicatures that arise from 'theme tunes' such as these. It is therefore problematic for his theory that, as he himself admits (Steedman 2006*b*, appendix), the phonetic evidence for some of these thematic tonal markers (particularly L+H*) is in dispute. Moreover, we will see evidence below that these same sorts of implicature can arise from the *interaction* of different phonetic and linguistic signals, or from more general phonetic signals over whole phrases. Hence we believe the burden of proof is on Steedman to show that the phonetic evidence supports these utterances being analysed as themes, rather than as examples of more general subordinate relationships between clauses, such as those argued for in Rhetorical Structure Theory (Mann & Thompson 1988), as well as interactions with more general prosodic signals, e.g. of emphasis. For instance, these cases could be an example of the relationship between the nucleus, e.g. *Alan is a klutz*, and the satellite Evidence, e.g. *he's a good badminton player*. We will return to this discussion in our analysis of the results in Chapter 4, and in the examples in Chapter 7.

The essential difficulty with these theories is deciding in which direction the implicature is really going. Is it the overall tune which carries the affective meaning, or are the meanings really composed from the tones themselves? In music, variations on a tune are both recognisable as such and can evoke similar emotional responses without the tones involved being

said to be individually meaningful. Although for some examples the intended implicature is easily apparent, it is difficult to test whether these theories really scale up to account for a wide range of language. Further, they need to account for the broader phonetic evidence.

In fact, it has been shown the *interaction* of tune, linguistic signal and global variation in pitch gives rise to these sorts of connotations. In a series of experiments, Scherer, Ladd & Silverman (1984) and Ladd, Silverman, Tolkmitt, Bergmann & Scherer (1985) studied the interaction of question-type (*declarative* versus *wh-*), intonation contour (rising or falling) and overall mean $f0$ and voice quality in perceptions of affective stance of different utterances in German. They found that while certain types of affect are perceived primarily through global $f0$ and voice quality features, e.g. *arousal* (*relaxed* versus *impatient*); other types are identified by the interaction of all these signals. For instance, the *challenging* stance was discerned by a combination of falling intonation with *declarative questions*, regardless of $f0$ level. *politeness* and *agreeableness* were perceived through a combination of low mean $f0$ and the 'expected' intonation contour type for the question type (i.e. falling=*wh-question*, rising=*declarative question*).[22] For example, in the following exchange, the parent's declarative question in (2.78a) feels like a polite reminder, in (2.78b) it is a stern rebuke (cf. Sag & Liberman 1975):

(2.78) Mum, can I go to the movies tonight?

a. 
You've done your HOMEWORK?

b. 
You've done your HOMEWORK?

This sort of effect has also been shown by Grabe, Gussenhoven, Haan, Marsi & Post (1998) in Dutch. In their experiment, they found that speakers judged utterances to be more *friendly* and *polite* when the preaccentual pitch was opposite to the pitch of the first accent, i.e. a low prehead was more favourable with a high following accent than a low following accent, and vice versa with a high prehead. In other words, these affective connotations do arise through implicature from the tonal contour, but from the interaction of different parts of the contour, and from the contour and the words, rather than from the tones themselves. In particular, there seems to be a strong correlation between *expectedness* and *politeness*, something a strictly compositional approach cannot capture. In section 3.3.2, we will see

[22] Other studies have found high pitch correlates with *politeness*, Scherer et al. (1984) suggest this may be culturally determined, i.e. whether it is more polite to be confident (low pitch) or submissive (high pitch) (see further below).

such evidence would fall out straight-forwardly from a probabilistic approach to tonal meaning.

Ladd et al.'s (1985) study showed global *f*0 signalled *arousal* independently of intonational tune. This is the way global variation in pitch is standardly taken to interact with the meaning of tunes: modifying the realisation of tonal categories without obscuring their identity (Ladd 1996, p. 35). However, it has often been claimed that the 'meanings' conveyed by pitch variation at both the global and local level reflect 'biological', and therefore universal, imperatives (Pike 1945, Liberman 1975, Bolinger 1978, Ohala 1994, Gussenhoven 2002). The idea is that smaller (usually female) animals have smaller larynxes than larger animals and thus produce sounds with a higher pitch range. Human communication is affected by this *frequency code*, so we associate high pitch with 'feminine' connotations (e.g. subordination and submissiveness) and low pitch with 'masculine' connotations (dominance and aggression) (Ohala 1994, Gussenhoven 2002).[23] Liberman (1975, pp. 132-48) claims that intonation is an *ideophonic* system. That is, unlike with morphemes, the relationship between the signifier and the signified is not arbitrary. Rather than being referential, meanings are typically *metaphorical*, influenced by "universal considerations", though they may become grammaticalised so they no longer reflect these considerations.

This theory has been argued to explain the interpretation of intonation contours. So, for example, questions (being more uncertain), are associated with high peaks and high final rises; whereas statements (being more confident), are associated with lower peaks and low boundaries. As Ladd (1996, ch. 4) points out, question intonation in fact varies considerably cross-linguistically, e.g. final falls in *wh-questions* in Hungarian. However, proponents of the ideophonic approach would argue that this results from established grammaticalisation processes, similar to the divorcing of a lexical item from its original meaning to serve a grammatical function (see Gussenhoven 2002). The hypothesis still holds if question intonation is more likely to be associated with high pitch cross-linguistically. More generally, if the system is ideophonic, we should expect it to be unstable, as tonal variation will always be simultaneously perceived on both a grammatical and 'biological' level, affecting interpretation. We return to the implications of this for the meanings of tonal categories in section 3.3.2.

Once more, the impact of this work on our study depends on which of these accounts one believes. If pitch accents and boundary tones are directly manipulated to signal information structure, and this in turn leads to affective connotations through implicature, then the signalling and perception of affect are directly relevant for us. If affective connotations arise from interactions of tune and message, then they could still be relevant as indirect evidence

[23] Gussenhoven also advances two other biological constraints on intonation, the *effort code* and *production code*; which, while interesting, do not seem as pertinent to the discussion here.

of information structure. The more the signalling of affective meanings is divorced from the proposed signalling of information structure, the more dubitable accounts become that information structure is signalled by intonational tune in the first place. Even if we believe such 'meanings' arise from global variations in pitch, the idea that the system is ideophonic has consequences for our study. All variations in pitch claimed to signal intonation categories or prosodic prominence in general could simultaneously convey 'inherent' meanings associated with that pitch variation.

2.4 Overview and the Way Forward

In this chapter, we have laid out the basic properties of information structure, the mechanism used to signal how each entity, predication, etc. should be interpreted in relation to the existing discourse structure, i.e. referencing, updating and/or altering it. We have also shown difficulties with standard accounts as to how this is marked prosodically in English. It is usually claimed that pitch accents F(ocus)-mark syntactic nodes. Focus can 'project' to higher constituents on the basis of syntactic rules. However, we saw that there are many examples of both 'optional' and apparently obligatory accents within and outside focussed constituents, as well as cases where focus projection does not seem to be syntactically constrained. We saw that the 'contrastiveness' of focal accents can be explained within Alternative Semantics; but suggested that strong or emphatic accents might still be linked to *restricted* kontrast interpretations, i.e. a limited alternative set in the context. We saw how focus marking interacts with relative givenness: it is claimed given elements in relation to the current proposition are deaccented. Again, we showed that the association with accenting per se is problematic. Finally, we showed that information structure has two dimensions: the kontrast/background distinction; and theme/rheme units. The latter is closely related to prosodic phrasing. We discussed whether, and how, kontrast within theme is prosodically distinct from kontrast within rheme, e.g. L+H*(LH%) versus H*(LL%) respectively. We showed that this is often conflated with a claim for separate 'contrastive' accents (i.e. *restricted* kontrast). We suggested much of the confusion may have arisen because the prosodic effects of each type of 'contrast' was not considered separately.

As we can see, many of the uncertainties in the semantic account arise directly from different understandings of the prosodic facts. There is a general correspondence between kontrast and prominence. However, it is not clear if this is between kontrast and accenting, and, if so, whether it explains the distribution of all accents, and, if not, how non-focal accents are distinguished. Or if the correspondence is between kontrast and prominence, what this means. Can all degrees of prominence mark kontrast, only nuclear prominence, or

does prominence vary gradiently, signalling information status/accessibility, etc.? What is the role of pitch accent and boundary tone type, if any?

Further, most of this work looks at information structure in isolation, and does not account for the interaction of prosodic signals of both lower and higher levels of meaning. In section 2.1, we saw certain word classes are more likely to be made prominent. Syntactic structure also strongly constrains prosodic phrasing, either requiring or restricting breaks. It is unclear how this interacts with prominence and phrasing marking information structure. In the last section, we reviewed proposals which extend the role of the tonal marking of information structure to concepts such as 'mutually believed'/'polarity in relation to the common ground' and 'speaker/hearer oriented/supposed', giving rise to illocutionary and affective connotations. These theories assume intonational meaning arises compositionally from the meaning of individual tonal events. However, we saw that it can be hard to separate these out from the meaning of the overall tune. Further, we showed evidence for more general phonetic effects in signalling these meanings, difficult to reconcile with these accounts. This evidence is relevant to the claim that theme/rheme status is signalled by intonational tune. If illocutionary and affective connotations do not arise from these implicatures, the argument as to why information structure is thought to be signalled by tune in the first place is weakened. Even if such theories hold, global pitch variation was argued to be directly meaningful, impacting on the affective connotations of utterances.

In the next chapter we begin with a discussion of the full expressive potential of prosodic prominence and phrasing. Most of the theoretical work on the prosodic signalling of information structure seems to assume a fairly flat structure of accents within largely linear phrases, along with (in some cases) the intonational tune. However, we know that prosody has a structure internal to itself. We will see that this structure admits many more levels of gradation in both prominence and phrasing than most of these accounts assume. When the expressive power of this structure is taken into account, many of the puzzles and problematic cases laid out above disappear. Indeed, we claim this prominence and phrasing structure is sufficient to convey information structure, including the distinction between thematic and rhematic kontrast. The corollary of our argument is that intonational tune is much *less* important to signalling information structure than has previously been claimed. It follows from this that information structure itself is probably not as significant as thought to the implicature of illocutionary and affective 'meanings'. At the end of the next chapter we return to this question, and suggest some explanations as to where the intuitions behind the theories discussed above come from.

The other prong of our argument, developed in the next chapter, is how these structures are related. We will claim that the segmental string is 'mapped' onto prosodic structure,

and that this mapping is subject to probabilistic constraints, including information structural constraints. In this chapter we have begun to see why we might need to consider that this mapping to be probabilistic. Lower level semantic factors, including lexical class and syntax, also affect prominence and phrasing. Further, higher level semantic factors including affective connotations related to emphasis, are partly signalled through prominence. We have also seen phonetic factors constrain prosodic structure itself, including phrase length. In the next chapter we will use these to show why information structure is interpreted from prosodic structure, rather than being directly signalled by it.

We end this chapter by briefly looking at how the description of the tonal contour has come to be so central to explanations of intonational meaning, and conjecture that it arises from a general approach to the study of prosody which may now be less necessary than it once was. It has long been recognised that prosody is "a highly complex phenomenon, one in which physical features such as frequency, intensity and time, as well as their psychological counterparts (pitch, loudness and duration) and their interaction all play a part" (Cohen & 't Hart 1967, p. 177). In the face of this complexity, most prosodic research has concentrated on pitch so that "by this very reduction, a better understanding of ... prosody in general could be obtained". Along with this concentration, the 'Dutch school' (Cohen & 't Hart 1967, 't Hart & Cohen 1973, 't Hart & Collier 1975) established an approach to deriving 'linguistically distinct' intonation patterns which became widespread through the work of Pierrehumbert (1980, p. 59), i.e. the first approach below:

> One approach attacks the problem by attempting to deduce a system of phonological representation for intonation from observed features of F0 contours. After constructing such a system, the next step is to compare the usage of F0 patterns which are phonologically distinct. The contrasting approach is to begin by identifying intonation patterns which seem to convey the same or different nuances. The second step is to construct a phonology which gives the same underlying representation to contours with the same meaning, and different representations to contours with different meanings.

Most of the subsequent work in the Pierrehumbert tradition adopted this philosophy, i.e. establish phonological categories first, attach meanings afterwards. 'Perceptually relevant' pitch movements were found experimentally to establish phonological categories. The problem with this is, unlike with segmental phonology, there is no clear basis on which this phonology lies. We do not know how subjects make 'perceptual relevance' judgements. Language is inherently a communicative medium, therefore our perception of distinctive units stems from our perception of meaningful units (see further in section 3.3.2). We do know, however, as we saw in section 2.3.3, that intonational meanings themselves are complex and can arise from the interaction of different signals at different levels of structure (e.g. *politeness*). Further, prosodic signals are complex: speakers can manipulate a limited number of

phonetic variables in a limited number of ways, to convey a huge range of meaningful distinctions. Therefore, similar cues are 'recycled' to convey (and perceive) different meanings at different levels of structure. The consequence of this is that it is very difficult to know, in the absence of context, what sorts of meaningful variation and how many prosodic cues are involved in one perceptible pitch variation.

We suggest something much closer to the second approach above may be possible if, in studying the prosodic correlates of any one type of meaning, we control for interacting effects of other levels of meaning as much as possible. This is not advocating a return to 'whole contour' theories which were rightly criticised (e.g. Sag & Liberman 1975). To do this with any degree of thoroughness has only recently become possible with the ability to computationally model multiple acoustic features in large collections of speech labelled for many linguistic features. Further, we can do this using the knowledge gained in the intervening years of prosodic research, particularly, that prosody has a structure in its own right. It places demonstrable constraints on its own realisation, both through articulatory constraints (cf. Mücke & Grice 2005, Xu 2005), and general pressures on organisation, e.g. rhythm and phrasing, as we show in the next chapter. Taking both sets of constraints in account (semantic and prosodic), it may be possible to identify the "intonation contours which seem to convey the same or different nuances"; and moreover, it might be ultimately both futile and uninteresting to try to identify abstract phonological intonation categories without regard to these constraints.

Chapter 3

How Prosody Conveys Information Structure

In the last chapter we described information structure, i.e. the basic organisation and salience of the information in an utterance in relation to the common ground of speakers in a conversation. In English, a primary cue to this structure is prosody; although there are many outstanding difficulties and uncertainties in standard theories describing the relationship between prosody and information structure. We saw at the beginning of the last chapter that much of the theoretical work on information structure describes its effect in isolation. However, lower level effects such as syntax and predictability also effect the same prosodic elements, i.e. pitch accenting and phrasing, which are claimed to signal information structure, compounding these difficulties. At the end of the chapter, we discussed theories which claim that basic information structure properties are signalled by intonational tune, i.e. tonal events, and that these information structure signals can then be manipulated to implicate higher illocutionary and affective connotations. However, despite a long history, these theories have not yet led to a generally agreed taxonomy of intonation events and meanings. It is also difficult to account for evidence of independent prosodic effects leading to the same connotations. The question is still open as to which direction the implicature can really be said to go; and on whether there really is definitive evidence that information structure is signalled by tonal pitch accent type. In this chapter, we put forward a quite different explanation for information structural phenomena within the framework of Autosegmental-Metrical (AM) prosody (term due to Ladd 1996). We will see that many of the uncertainties described above can be resolved when the full richness of expression within metrical prosodic structure is taken into account.

We begin by setting out the basic properties of the AM framework, which has become the established basis for prosodic description over the past 25 years. AM assumes a hi-

erarchical organisation of metrical prosodic prominence and phrasing. We will argue that phrasal organisation and relative prominence relationships within this structure are recursive, a property that is crucial to the argument we develop below. The framework also describes the intonational tune in terms of a linear series of intonational events, defined by (H)igh and (L)ow tonal targets that associate with prominences and boundaries at the phrase level. The standard annotation system for these events is To(nes) and B(reak) I(ndices) (Silverman et al. 1992, Beckman & Hirschberg 1999), which we will describe. Although we accept the basic principles of this description, we argue that tune is much less important to the description of information structure than is usually supposed. Variation in pitch range at the phrase level is assumed to be *para-linguistic* and *gradient* (see summary in Ladd 1996, ch. 1), though we argue that interaction with pitch variation signalling prominence structure needs to be carefully accounted for.

In the next section we advance our theory of how prosody signals information structure. We argue that the basic correspondence is between salience (or kontrast) and nuclear accenting, and between organisation and phrasing. We argue that these relationships should be conceived as probabilistic constraints because of the other established influences on each structure, as well as their relationship with each other. This can lead to ambiguity in the information structure interpretation of any given prosodic structure, but we argue this is mediated by the likelihood of each interpretation given that structure controlling for other constraints. We claim that information structure is not signalled by intonation type, but rather that the theme/rheme division is signalled by relative prominence above the phrase level. We end with a discussion of the status of emphatic accents, which we claim signal *restricted* kontrast, as well as a note about relevant phonetic features in our inquiry.

This leads us to a discussion about the nature of prosodic units in general. Drawing from recent work on probabilistic language processing in other fields, we extend our argument about why the relationship between information and prosodic structure should be seen as probabilistic. We will also look at the notion of *markedness* in conveying meaning prosodically given the discussion in the first two sections. Finally we return to the debate at the end of the last chapter about how tune conveys meaning, i.e. tones or tunes. We consider whether this is really an all-or-nothing question given recent work questioning the compositional nature of meaning and categoriality in general. This impacts on the question of whether 'gradient' prosodic variation really forms a separate perceptual stream, and does not impact significantly on our perception of prosodic categories. We end the chapter by setting out our approach to testing the predictions of our theory in the rest of the thesis.

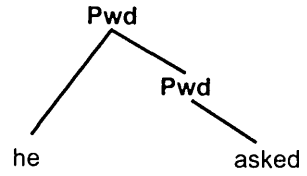


Figure 3.1: Integration of a function word into a recursive prosodic word (adapted from Shattuck-Hufnagel & Turk 1996, p. 217).

3.1 Prosodic Concepts

This section introduces the elements of the AM model of prosody which we will assume in the rest of the thesis. Basic properties of the model that are now widely accepted in the literature will be set out briefly. On more contentious issues, we offer argument and evidence for our viewpoint. We will see that, according to this model, prosody is defined by two basic components: a hierarchical structure of prosodic phrases and prominence; and intonation events which associate with phrase boundaries and prominences, to describe the *tune* of the phrase. In general, global variation in $f0$ range forms a separate 'stream' of information, which influences the way the basic prosodic elements are realised, without obscuring their structure. However, we will show that this variation needs to be treated carefully, as modificaton of $f0$ range can also be used to signal distinctions internal to prosodic structure.

3.1.1 Phrasing

Prosodic structure is built from a nested hierarchy of constituent elements. It is generally agreed that there are a limited set of such elements, each with particular characteristics (e.g. Prince & Liberman 1977, Beckman & Pierrehumbert 1986, Nespor & Vogel 1986, Gussenhoven 1988, Hayes 1989, Wightman, Shattuck-Hufnagel, Ostendorf & Price 1992, Ladd 1996, Shattuck-Hufnagel & Turk 1996). However, there is less agreement on the exact inventory, and on the restrictions which govern the vertical combination of these constituents, i.e. strictly layered (e.g. Hayes 1984) versus recursive (see discussion in Shattuck-Hufnagel & Turk 1996). Here, we take the view that the structure is, in principle, infinitely recursive, with a performance-based limit on the depth of structure found in natural language (as per Ladd 1996, ch. 6). Below we briefly set out these basic constituents and some evidence for recursive phrasing structure.

The basic prosodic unit is the *syllable*, which minimally defines higher prosodic groupings (see Shattuck-Hufnagel & Turk 1996, p. 219). We adopt one prosodic phrasing unit between the syllable and the *phonological phrase* (PhP), the *prosodic word* (Pwd) (after Shattuck-Hufnagel & Turk 1996).[1] At this level, we begin to see the advantage of allowing for recursive phrasal structure. It is often noted that content words and adjacent function words or pronouns act as a single unit prosodically. Studies show elision and contraction effects within but not across prosodic word boundaries, e.g. *wanna* contraction, or greater pausing and word-initial lengthening between *John* and *asked* in *John asked* than between *he* and *asked* in *he asked* (Grosjean, Grosjean & Lane 1979, Gee & Grosjean 1983, Turk & Shattuck-Hufnagel 2000). Unlike other words, function word duration is affected by its predictability given the following word, suggesting a single processing unit (Bell et al. 2003). However, orthographically and semantically, in such cases there are still two separate units, something which can be neatly and straight-forwardly represented by recursive structure, as in Figure 3.1.

The basic unit of prosodic phrasing is the *phonological phrase* (PhP) (term re Nespor & Vogel (1986), Hayes (1989) and Shattuck-Hufnagel & Turk (1996)).[2] This phrasing serves a 'chunking' function, breaking the continuous speech stream into units to aid both production and perception (e.g. see van Wijk's (1987) analysis of Gee & Grosjean (1983)). PhP boundaries are correlated with well documented phonetic cues including initial strengthening (i.e. strong articulation of the first sound), pre-boundary lengthening (i.e. greatly increased duration in the last few syllables), pausing, pitch movement associated with the boundary, and the blocking of elision and contractions (all cited in the discussion in Shattuck-Hufnagel & Turk (1996), see also Warren (1999, pp. 166-7)). Phrasing is constrained by eurhythmic effects: Grosjean et al. (1979) and Gee & Grosjean (1983) show that speakers place boundaries in the middle of syntactic units in order to keep the sizes of prosodic units approximately equal. Further, there is the general tendency of speakers to 'prosodify' lists, or other groups of words like telephone numbers, that have no particular syntactic structure (Suci 1967). As will see below, the PhP is also the smallest domain of the effects of declination (the gradual reduction in $f0$ levels), and downstep (phonologically significant pitch accent height reduction).

There is broad agreement that PhPs can be grouped together to form higher levels of structure. In many descriptions these groupings are taken to be from a limited number of dif-

[1] Different theories argue for various combinations of constituent types in this region of the structure (see discussion Shattuck-Hufnagel & Turk 1996, pp. 215-9). However, the disagreements involved are for the most part orthogonal to our purposes. Our notion of the prosodic word includes the functions argued for Nespor & Vogel's (1986) and Hayes's (1989) *clitic group* and Selkirk's (1995) *minor phrase*. It also follows from Beckman & Pierrehumbert's (1986) discussion that there is little evidence for a separate *accentual phrase* in English, and that therefore the *prosodic word* is the only grouping between the syllable and the PhP.

[2] It is equivalent to Beckman & Pierrehumbert's (1986) *intermediate intonation phrase*, and Selkirk's (1995) *major phrase*.

ferent constituent types, most commonly the *intonation phrase* (IP) (Nespor & Vogel 1986, Hayes 1989, Beckman & Pierrehumbert 1986, Selkirk 1995, Shattuck-Hufnagel & Turk 1996), as well as larger groupings such as the *utterance* (Nespor & Vogel 1986, Hayes 1989, Selkirk 1995) or *paragraph* (Hirst & Cristo 1999). Phonetic evidence for these groupings includes more consistent or enhanced use of cues to phrase breaks listed above (Wightman et al. 1992, Chavarría, Yoon, Cole & Hasegawa-Johnson 2004, Redi & Shattuck-Hufnagel 2001); as well as declination effects over the larger phrasal unit (e.g. de Pijper & Sanderman 1994, Swerts 1997).

However, the main evidence for qualitatively different levels of phrasing above the PhP is usually association with different levels of syntactic and discourse boundaries, e.g. syntactic constituents versus clauses (see discussion in Shattuck-Hufnagel & Turk 1996). As we discussed in section 2.1, such syntactic distinctions between prosodic phrasing type have always been difficult to reconcile with the apparent latitude speakers have to use different prosodic groupings to convey the same syntax; as well as the frequent mismatches between phrase level (i.e. PhP versus IP) and syntax structure level. For example, a particularly deliberate or emphatic rendition of the following could have phrase boundaries after every adjective phrase:

(3.1) A: What did you want?

B: (a PALE)$_{PhP}$ (ORANGE)$_{PhP}$ (and YELLOW)$_{PhP}$ (BALLGOWN)$_{PhP}$!

Further, there has never been any general agreement on the inventory of higher-level phrases needed to convey hierarchical clause and discourse structure.

The difficulty comes from trying to reconcile a non-recursive hierarchy of prosodic phrase types (i.e. the Strict Layer Hypothesis) with a patently recursive syntax and discourse structure (cf. Ladd 1996, ch. 6). If one takes the basic PhP as in principle recursive, the syntactic attachment difficulties (e.g. (3.1)) fall out easily. Syntactic boundaries are marked by prosodic boundaries *at some level* of prosodic structure (Cooper & Paccia-Cooper 1980, Ladd 1986, Ladd 1996, Truckenbrodt 1999, Wagner 2003). So, the phrases in (3.1) would all be grouped in one higher PhP. Hierarchical clause and discourse structure is immediately accounted for by the same mechanism. This includes, as noted in section 2.2.3.1, the disagreement as to the appropriate level at which to define theme and rheme units. We can now state that these are marked by prosodic phrases at some level of phrasing structure. Of course, in practical terms, there is a limit to the number of degrees of boundary strength that can be reliably distinguished on the basis of phonetic cues. For example, Wightman et al. (1992) and Ladd & Campbell (1991) report listeners can reliably distinguish four levels above the prosodic word. de Pijper & Sanderman (1994) showed subjects can reliably use a ten-point

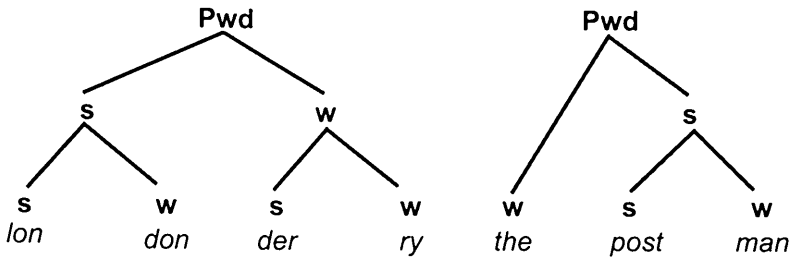


Figure 3.2: Metrical structure of the words *Londonderry* and *the postman*.

scale from word-level up. However, under our view, the boundary strength index is merely an annotational convention, which may be correlated with the strength of syntactic/discourse boundaries, but is not identified by them. In this study, therefore, we take the basic unit of phrasing to be the PhP. Higher groupings are recursive applications of the same basic phrase, not qualitatively different types. We will see below that this view is corroborated by pitch scaling effects across boundaries (see further Ladd 1996, ch. 6).

3.1.2 Prominence

Syllables are also the basic unit on which prominence relationships are defined. In AM theory, syllables map onto a hierarchically organised metrical structure, i.e. a binary branching structure of *w(eak)* and *s(trong)* nodes (Liberman 1975). This structure creates relative prominence relationships between phrasal constituents, i.e. syllables within a Pwd, and then in turn Pwds with a PhP, and among PhPs. Each prosodic word is represented by a branching structure containing at least one *s* node, see Figure 3.2. Again function words can form part of the same unit in these structures (cf. Turk 1999). The node in each word which is only dominated by other *s* nodes, the nucleus, is usually the attachment point to higher levels of structure (although other *s* nodes may take part in stress shift). Prominence relationships at the phrase level are then formed in the same manner.

The pattern of relative prominence creates the perception of *rhythm*, which is fundamental to human language processing. It is one of the earliest properties of speech infants attend to (Nazzi & Ramus 2003), perhaps related to the greater human perceptual sensitivity to change than absolute values (Kluender, Coady & Kiefte 2003). In English at least, rhythm in turn constrains metrical structure, i.e. there should be approximate *perceptual isochrony* been equal prominences at each level of the structure (e.g. Huss 1978, Terken &

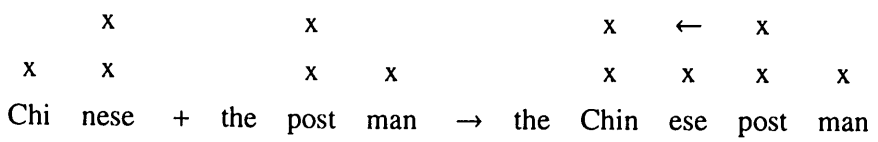


Figure 3.3: Effect of stress clash on the metrical structure of *the Chinese postman*.

Hermes 2000). Although this isochrony is not absolute, the effects of *stress clash* are evident (Prince & Liberman 1977). That is if two strong beats occur together, one stress is 'moved' to avoid the clash, see Figure 3.3 (using metrical grid notation from Liberman 1975). Or if there are too many weak beats together at one level, an extra beat is 'added' to the next strongest node to preserve the rhythm (Halle & Vergnaud 1987, Hayes 1995). Rhythmic disruption causes difficulty in speech perception (see review in Cutler et al. 1997). Experimental and corpus-based work has confirmed these effects and shown them to be widespread (Shattuck-Hufnagel, Ostendorf & Ross 1994, Grabe & Warren 1995, Harrington, Beckman, Fletcher & Palethorp 1998, Ramus, Nespor & Mehler 1999, Grabe & Low 2002).

Importantly, both words and metrical nodes form independent structures, which are then 'mapped' onto each other (cf. Liberman 1975). This mapping is constrained by the properties of each structure, as well as constraints on their relationship. In English there is a strong constraint aligning lexically stressed syllables with strong beats, so weak syllables can be 'squished together' between strong nodes in the final output (Halliday 1967, Liberman 1975).

Prominence is correlated with a variety of phonetic cues, including vowel quality, i.e. *reduced* or *unreduced*,[3] increased duration, intensity and spectral tilt (see Terken & Hermes 2000, Kochanski, Grabe, Coleman & Rosner 2005). At the phrase level, it is also correlated with pitch accenting, i.e. a localised pitch movement. According to many accounts (the *accent-first* theory of phrasal stress (Selkirk 1984)) the former are cues to lexical stress, while pitch accents attach to lexically stressed syllables to mark phrase level prominence. However, as discussed in section 2.1, different studies have found conflicting evidence about whether lexical stress is consistently marked in unaccented positions. Further, as we will see below and in our corpus study, listeners often hear a definite prominence at the phrase level when there is little or no pitch movement.

Again, this evidence can be straight-forwardly explained if we take prominence to be a property of recursive structure (the *stress-first* theory (Ladd 1996, ch. 6)). Increasing acoustic prominence is correlated with increasing levels of relative prominence, and pitch

[3]Note that this distinction may not be categorical (Fear et al. 1995).

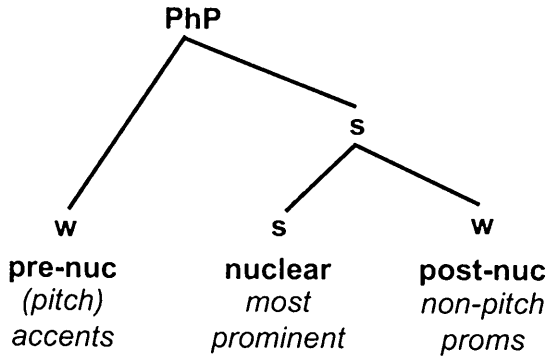


Figure 3.4: Basic phrase level metrical structure.

movement is one of the markers of prominence at the phrase level. Therefore we would expect lexically stressed syllables to have varying levels of acoustic prominence depending on their place in this structure, and pitch movement to interact with other cues to phrase level prominence. In fact, increased duration may be necessary for the perception of an accent (see Dilley & Brown 2005, pp. 32-4); certainly the presence of an accent lengthens all the syllables in a prosodic word (Turk & Sawusch 1997, Turk 1999, Cambier-Langeveld & Turk 1999). Recent evidence has suggested loudness may be a more salient cue to phrase-level prominence than $f0$ level or movement (see Kochanski et al. 2005). In this work, we will continue to refer to phrase level prominence as *accenting*, as it is such a widely used and convenient term. However, we should be very clear that accenting does not necessary imply pitch movement, rather it encodes the perception of phrase level prominence (see further in section 3.2.5). The prominence of a syllable is also crucially perceived relative to metrical structure, so we do not expect a direct relationship between the perceived degree of prominence and the phonetic properties of that syllable (Ladd 1996). Like in music, once a rhythm has been established, the *expectation* of a strong beat may be enough for people to perceive one, even without discernible phonetic cues. We will see that the recognition that prominence is *relational* is very important to the theory we develop below.

In particular, the perception of *nuclear prominence*, which is fundamental to our theory, arises directly from metrical structure. The nuclear accent falls on the word at the PhP level dominated entirely by strong nodes. By default this structure is right-branching, so the nuclear accent is usually the right-most in a phrase and is perceived as the most *structurally* prominent (see Halliday 1970). Basic phrase structure is therefore as in Figure 3.4 (cf.

Liberman 1975). Strong nodes to the left of the nuclear accent can carry *pre-nuclear accents*, which may be marked by pitch movement. Strong nodes to the right may be accented, but, in English at least, are much less acoustically prominent, and are marked by very little or no pitch movement (cf. Grice, Ladd & Arvaniti 2000).[4]

The perception of nuclear accenting cannot be reduced to the relative height of accents, as listeners can both perceive a high early accent to be higher *and* a late low accent (i.e. downstepped) to be more structurally prominent, that is consistent with broad or object-focus (Rump & Collier 1996). Further, Ayers (1996) has shown that response times in a phoneme-monitoring task were slower for downstepped nuclear accents than other nuclear accents (suggesting they are less phonetically prominent), but the type of nuclear accent (downstepped versus not) did not affect times in a question-answering task. These results cannot be explained by *declination*, i.e. well-documented evidence that pitch peaks are perceived to be higher the later in a phrase they appear (e.g. Rietveld & Gussenhoven 1985, Gussenhoven & Rietveld 1988, Terken 1991); as it has been shown that even taking this into account, speakers have a rightward bias (see early work in the 'nuclear tone' tradition (O'Connor & Arnold 1961, Crystal 1969), and later experimental work (Rump & Collier 1996, Terken & Hermes 2000)). Further, given an utterance with no pitch movement, listeners 'hear' an accent on the right-most stressed syllable; their *expectation* is sufficient to perceive nuclear prominence (Hermes & Rump 1994). This perception holds even when the accent is obscured by noise (Xu, Ching & Xuejing 2004).[5]

Nuclear accents are often followed by an L- phrase accent, which lowers the post-nuclear pitch range (Beckman & Pierrehumbert 1986, Beckman 1996). This could be argued to lead to the perception of nuclear prominence because the nuclear accent 'stands out' more. In fact, Xu & Xu (2005) claims that pitch range is directly manipulated to express focus, so the nuclear accent is simply the last in the focal region, similar to Japanese (Beckman & Pierrehumbert 1986, Sugahara 2003), or Mandarin (Xu 1999). However, this does not explain the perception of nuclear prominence in downstepped final accents with no significant fall (cf. Ayers 1996). Nor, as we shall see in section 3.2, does Xu & Xu's (2005) account cover the range of prosodic focus effects in English that we are interested in. This issue is difficult because nuclear accents are often followed by a fall in pitch (consistent with an L tone) (see Grice et al. 2000); but since this is not necessary for their perception, we maintain it is determined by metrical structure.

[4]Note that Grice et al. (2000) analyse these 'accents' as being associated with phrase tones. They cite evidence from other languages showing they may therefore be either high or low (H- versus L-), and may be as acoustically prominent as the nuclear accent. Nevertheless, they are perceived in these languages as semantically subordinate to the nuclear accent, hence their post-nuclear status.

[5]These authors view this in terms of the perception of focus position, however they assume focus is always realised on the nuclear accent (in our terms).

There is some evidence, however, for consistent phonetic and distributional differences between nuclear and pre-nuclear accents. Schepman, Lickley & Ladd (2006) have shown that the peak in pre-nuclear accents is consistently aligned later than the peak in nuclear accents in Dutch; their recent work has confirmed the effect in English (Ladd, Schepman, White, Quarmby & Stackhouse in preparation).[6] Ayers (1996) showed that nuclear accents, but not pre-nuclear accents, improved response times in both phoneme monitoring and question-answering tasks. Finally, pre-nuclear accents are much more susceptible to stress shift, i.e. movement to an earlier strong node because of stress clash, than nuclear accents, which is exactly what we would expect if the position of pre-nuclear accents was determined by metrical structure (Shattuck-Hufnagel et al. 1994).

One potential complication, given the centrality of nuclear prominence to our theory is that, although there is reasonable agreement between theorists and experimental subjects as to degrees of boundary strength on phonetic grounds (see above), the resulting phrases do not always accord with our expectations on structural grounds (i.e. re Figure 3.4), as Ladd (1996, pp.235-51) points out. Ladd identifies three such cases. The first is when a phrase seems ill-formed because of a disfluency or false start (cf. Brown, Currie & Kenworthy 1980). Here we simply claim these utterances *are* ill-formed, though of course there needs to be independent motivation for identifying disfluencies. The second is when a phrase that seems structurally complete does not have clear phonetic boundary cues. Anecdotally, this is a feature of rapid speech, when all durational cues are weakened, and the structural ambiguity can be seen as a feature of such speaking styles (e.g. see Beckman 1996, pp. 54-7).

The last is more problematic, and brings us back to the arguments for recursive phrase structure above. In (3.1), we want to say that each adjective phrase forms its own PhP, *and* that *ballgown* is nuclear. If relative prominence relationships can de defined above the phrase level, in line with recursive phrasing, this falls out straight-forwardly. The last in a series of roughly equally acoustically prominent nuclear accents will be perceived as the most structurally prominent over the larger phrase, e.g.:

(3.2) I ordered...

(												*			)*PhP*
(	*										)*PhP*	(	*		)*PhP*
(	*	)*PhP*	(	*	)*PhP*	(		*		)*PhP*	(	*			)*PhP*
(a pale		)*Pwd*	(orange		)*Pwd*	(and		yellow		)*Pwd*	(ballgown				)*Pwd*

[6] In Silverman & Pierrehumbert's (1990) influential study, they claim to show there is no difference in the alignment of pre-nuclear and nuclear peaks, whereas in fact they show that nuclear peaks are earlier (p. 96). They claim this is because of the influence of the following L- phrase accent. While this may be part of the story, we would say this is a possible explanation of the effect (re the discussion above), rather than evidence the effect does not exist.

That is, prominence relationships mirror that at the phrase level. For a phrase to be perceived as post-nuclear, its nuclear accent must be significantly less acoustically prominent than the preceding nuclear accent. We will see in section 3.2.3 that the recognition of this property of metrical structure is crucial in our explanation of the signalling of information structure.

3.1.3 Intonation Events

The other main component of the AM model is the intonational tune, or melody. In AM theory, this is composed from a series of tonal events: pitch accents and edge tones, which are associated with phrase-level prominences and phrase boundaries respectively. The current standard annotation system to describe these tones is ToBI. We will use this standard here, with modifications discussed below. The ToBI system (Silverman et al. 1992) is largely drawn from the work of Pierrehumbert (Pierrehumbert 1980, Beckman & Pierrehumbert 1986). We will concentrate on the current standard for American English (Beckman & Hirschberg 1999), though we draw from the earlier works where relevant.

Phrase boundary strength is indicated in terms of a break index value after each word boundary. 0 and 1 are used for word boundaries. 2 is used for 'mismatches' between the tune structure and phonetic cues to juncture. 3 is used for PhP boundaries and 4 for IP boundaries. Intonation events are defined in terms of H(igh) and L(ow) tonal targets, i.e. target points in the pitch span or 'tonal space' (re Ladd 1996, p. 73) of the current phrase. Pitch accents are described by a starred H* or L* associated with the stressed syllable, plus an optional L or H target immediately preceding or following it. Of the resulting six logical possibilities, L*, H*, L+H* and L*+H are held to be part of the standard description of English (H*+L and H+L* were included in earlier versions (Beckman & Pierrehumbert 1986), combinations of identical targets are not allowed). In addition, accents involving an H target can be described as downstepped (using !) if they are phonologically lower than a preceding H target, thereby adding !H*, L+!H*, L*+!H*, H+!H* (the latter a clear step down to the accented syllable from a H target which cannot otherwise be accounted for). We will discuss downstepping further below. There are two types of edge tones. Phrase accents (L- and H-) mark PhP boundaries (at break level 3). They are said to describe the behaviour of the $f0$ curve from after the last accent to the end of the phrase, i.e. low or rising. Boundary tones (L% and H%), on the other hand, only occur at IP boundaries and describe a local rising or falling $f0$ movement associated with the boundary. Examples of some of the different combinations can be seen in Figure 3.5.

In the 15 years since its consolidation as an annotation system, there have been several major studies reporting annotator agreement using ToBI on a variety of corpora (Silverman

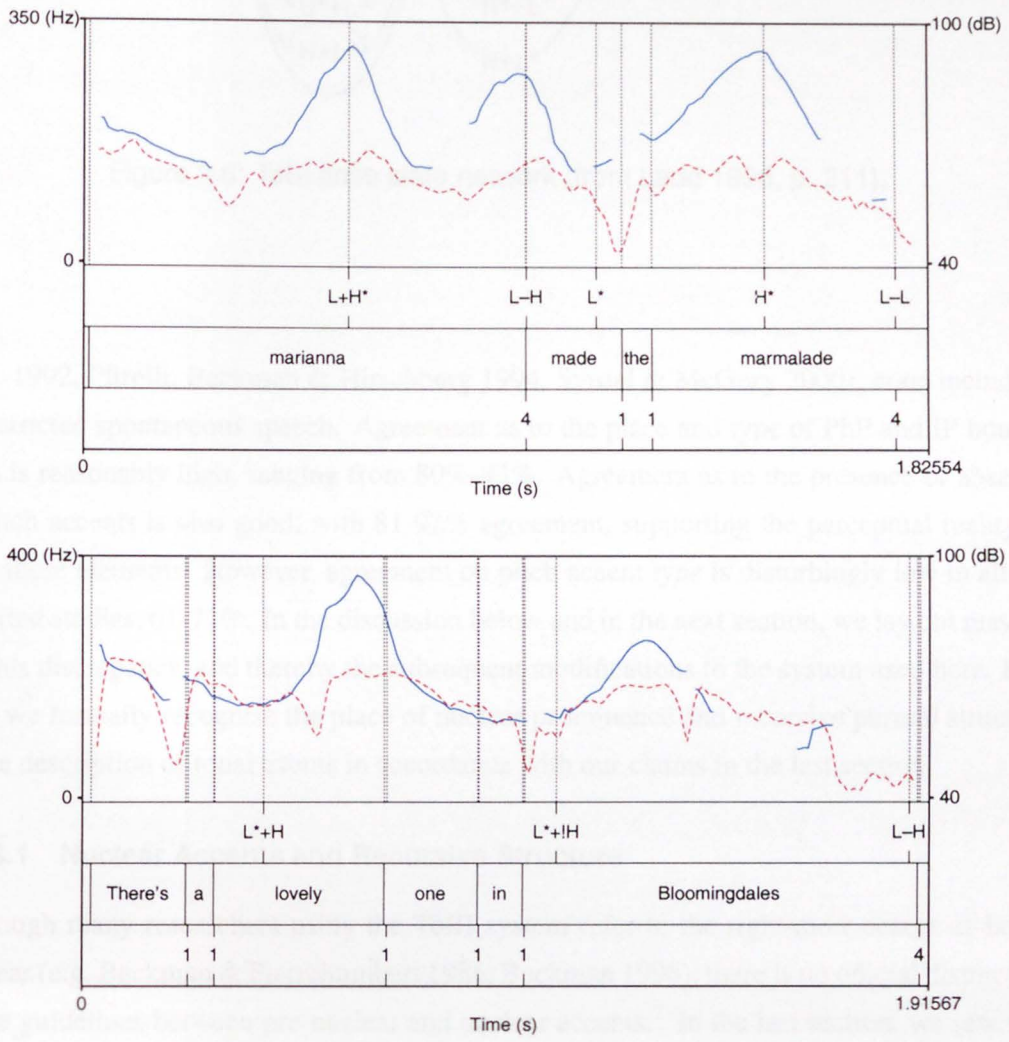


Figure 3.5: Examples of tones and break indices, from ToBI annotation guidelines (Beckman & Elam 1997) ($f0$ trace is the blue line and intensity curve the dashed red line).

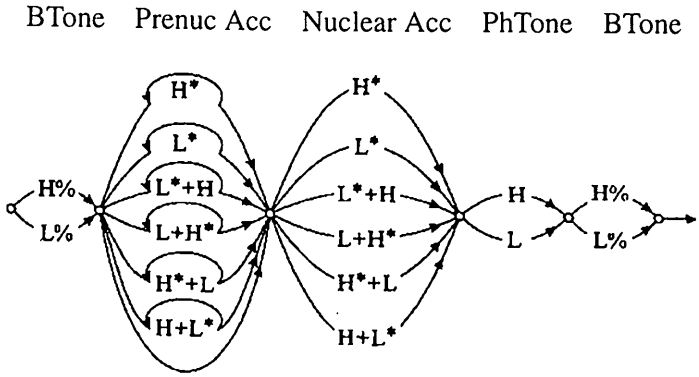


Figure 3.6: ToBI finite state network (from Ladd 1996, p. 211).

et al. 1992, Pitrelli, Beckman & Hirschberg 1994, Syrdal & McGory 2000), none including unrestricted spontaneous speech. Agreement as to the place and type of PhP and IP boundaries is reasonably high, ranging from 80%-93%. Agreement as to the presence or absence of pitch accents is also good, with 81-92% agreement, supporting the perceptual reality of both these elements. However, agreement on pitch accent *type* is disturbingly low in all the reported studies, 61-72%. In the discussion below and in the next section, we lay out reasons for this discrepancy, and thereby the subsequent modifications to the system used here. Further, we formally recognise the place of nuclear prominence and recursive phrasal structure in the description of tonal events in accordance with our claims in the last section.

3.1.3.1 Nuclear Accents and Recursive Structure

Although many researchers using the ToBI system refer to the right-most accent as being nuclear (e.g. Beckman & Pierrehumbert 1986, Beckman 1996), there is no official distinction in the guidelines between pre-nuclear and nuclear accents.[7] In the last section, we saw that the nuclear accent is the 'perceptual centre' of the phrase, therefore it would not be surprising that its tonal information should have special status. As argued by Ladd (1996, pp. 206-211), intonation tunes with arbitrary numbers of pre-nuclear accents, but the same nuclear accent tone type tend to be perceived as the 'same' contour, an idea standard in the 'nuclear tone' tradition (O'Connor & Arnold 1961, Crystal 1969). Here we follow Ladd in arguing that this idea is not incompatible with AM theory, given a minor modification to Pierrehumbert's (1980) finite state grammar for the composition of intonational tunes, shown in Figure 3.6. We will see in section 3.3.2 that this position is corroborated by tone type distributional

[7] Although this has recently been discussed (Beckman et al. 2005).

evidence. In fact, as our later experimental and corpus work suggests, it may be that the tonal accent in the nucleus 'characterises' the phrase, i.e. it more consistently conveys the illocutionary connotations associated with different tonal types than accents in general.

We also argued above that there are no qualitatively distinct phrase types above the PhP. The ToBI system, however, assumes a two-way distinction between PhPs and IPs, with boundary tones only associated with the latter. Although it is not crucial to our argument, we would suggest that, rather than a categorical distinction, larger phrasal groupings are more likely to be associated with full boundary tone movements than smaller ones.

3.1.3.2 Tone Alignment

In ToBI, accents are described as 'associated' with stressed syllables, i.e. the H (in H* and L+H*) or L (in L* and L*+H) tonal target is *perceived* as falling on the stressed syllable. A growing body of work has shown that, at least in carefully controlled laboratory conditions, these targets are in fact closely aligned with syllable offsets and onsets (Arvaniti, Ladd & Mennen 1998, Ladd, Mennen & Schepman 2000, Ladd & Schepman 2003, Atterer & Ladd 2004, Ladd 2004, Xu & Xu 2005) under changes in speaking rate. It has been suggested this is due to a phasing relationship between $f0$ fluctuations and articulatory movements marking the beginnings and ends of syllables (Xu 2005, Mücke & Grice 2005). This raises the intriguing possibility that the categorical nature of tonal alignment (i.e. by syllable on/offset, rather than gradient) may result from articulatory pressures on the timing of $f0$ peaks.

The only exception to this is the distinction between L+H* and H*, which both have their peak in the stressed syllable (see Figure 3.5). Supposedly, L+H* is distinguished by a preceding L target, though most H* accents appear to have this target as well (Ladd, Faulkner, Faulkner & Schepman 1999, Ladd & Schepman 2003, Xu & Xu 2005). Tellingly, this distinction causes major difficulties for annotators: Silverman et al. (1992) and Pitrelli et al. (1994) collapse the categories and do not report agreement at all. Syrdal & McGory (2000) say that it is the most common cause of pitch accent type disagreement. However, we saw in section 2.2.3 that L+H* and H* have been argued to have distinct and important functions in information structure. In the next chapter, we review the phonetic and semantic evidence in detail, and conclude the functional distinction lies elsewhere in the prosodic system, and that there seems to be no basis for a categorical distinction between L+H* and H*.

Apart from this, the theoretical definition of pitch accent types solely in terms of tonal alignment is problematic. The ToBI guidelines talk about "an apparent tonal target on the accented syllable" (Beckman & Hirschberg 1999), rather than alignment because, in different contexts, peaks and valleys quite regularly fall after, and sometimes before, the stressed

syllable. Many studies show systematic effects on target location that are as large or larger than those attributed to categorial shift (see review in Wichmann, House & Rietveld 2000). For instance, van Santen & Möbius (2000) found significant variation due to segmental structure in H*LL% contours, e.g. peak location is systematically later in sonorant-final accent groups than in obstruent-final accent groups (e.g. *pin* versus *pit*). A similar peak delay has been found between phonologically long and short vowels (Ladd (2004, p. 125), Xu & Xu (2005)); as well as large effects due to the location of the stressed syllable in polysyllabic words (Silverman & Pierrehumbert 1990). 'Tonal crowding', i.e. from following tones or prosodic boundaries, can also cause leftward shift in the $f0$ peak (Silverman & Pierrehumbert 1990, Wichmann et al. 2000, Arvaniti, Ladd & Mennen 2006). Wichmann et al. (2000) show $f0$ peaks on accents at the beginning of topic-initial sentences are consistently later than in non-initial sentences. Emotional state may also impact target alignment independently of $f0$ level. For example, Bänziger & Scherer (2005) found emotions with 'high arousal' (such as *elation* and *hot anger*), had slightly steeper rises and steeper falls, i.e. L and H targets closer together, than emotions with 'low arousal' (such as *sadness*, *happiness* or *cold anger*), controlling for $f0$ level.

Now, co-articulation and neutralisation effects are frequent in the realisation of ordinary phones in different contexts, so this does not invalidate the existence of these accent types. However, it does question the value of these alignments as the sole, or even primary, cue to pitch accent type. For instance, Pierrehumbert & Steele (1989) claim to have found evidence for a categorical perception boundary between L*+H and L+H* in an imitation task, using the stimulus *Only a MILlionaire* (see examples of these accents in Figure 3.5). However, recently Shattuck-Hufnagel, Dilley, Veilleux, Brugos & Speer (2004) showed in a similar experiment that the result could be explained by stress shift from *mil'lionaire* to *million'aire* rather than tonal type. Nothing rules out this explanation. There has been surprisingly little experimental work testing whether *meaning differences* associated with the different accents in English are in fact cued primarily by tonal alignment.

Given this evidence, we suggest the meaning differences attributed to ToBI pitch accents (as we saw in sections 2.3.2 and 2.3.3) are in fact signalled by multiple phonetic cues, potentially at different levels of prosodic structure. Tonal alignment can be *exploited* to achieve an interpretative boundary, but it is not necessarily *perceived* as one (see further in section 3.3). In fact, speech synthesis systems wishing to generate pitch accents types, based on either ToBI-like categories or functional ones (e.g. question, continuation), routinely use a large range of phonetic features; including specifications for multiple points in the accent, as well as mean $f0$, intensity and duration (van Santen & Möbius 2000, Taylor 2000, Pan et al. 2002, Clark 2003).

3.1.3.3 ToBI and Intonational Meaning

This brings us back to the debate at the end of the last chapter, and that is the relationship between ToBI events and intonational meaning. We should be clear that ToBI in itself is an annotation system, and only claims to be able to describe phonologically distinct variations in intonational tune. It is compatible both with a strictly compositional approach to intonational meaning, and 'whole contour' theories (e.g. see discussion of Fujisaki (1981) in Liberman & Pierrehumbert (1984)). However, as we saw in the last chapter, many of its leading proponents have used ToBI in conjunction with 'accent first' theories of phrasal stress, which for the reasons discussed above we believe is misguided. Further, we believe the concentration on ToBI description has distracted from the importance of relative prominence and phrasing in signalling intonational meaning. A major theme in this thesis is to recognise how much of the basic organisation of information in English is signalled by metrical structure; and to therefore put intonational tune in its proper place in prosodic description.

These influential theories, assuming the 'accent first' view of phrasal stress, have argued for a strictly compositional approach to intonational meaning, i.e. each tonal event has a meaning, and these events join together compositionally to derive the prosodic meaning of the phrase (e.g. Pierrehumbert & Hirschberg 1990, Steedman 2000, Steedman 2006*a*, Steedman 2006*b*). However, there is still little agreement on the 'meanings' of particular pitch accents and boundary tones. Further, as we discussed in the last chapter, relevant proposals are hard to pin down to verifiable claims about the meanings of particular utterances. We will look at this more closely in section 3.3, suggesting a hybrid approach between compositional and whole contour analyses.

More importantly for our purposes, since ToBI describes intonational tune, it does not annotate any levels of prominence other than 'accented', which is assumed to be associated with lexical stress. This may have led many researchers to begin by assuming all categorial distinctions between accents can be framed in terms of sequences of tones, and therefore that all meaning distinctions should be thus described. However, as is pointed out by Taylor (2000), in reported studies the vast majority of accents are H* or L+H*: they comprise 94% in the Boston Radio News corpus; Syrdal & McGory (2000) and Pitrelli (2004) report 90% and 83% respectively in corpora of professionally read speech, Hedberg & Sosa (2001) found 81% in televised political talk shows. Evidently much information - we argue particularly that relevant to the signalling of information structure - is carried by variation in the realisation of these accents.

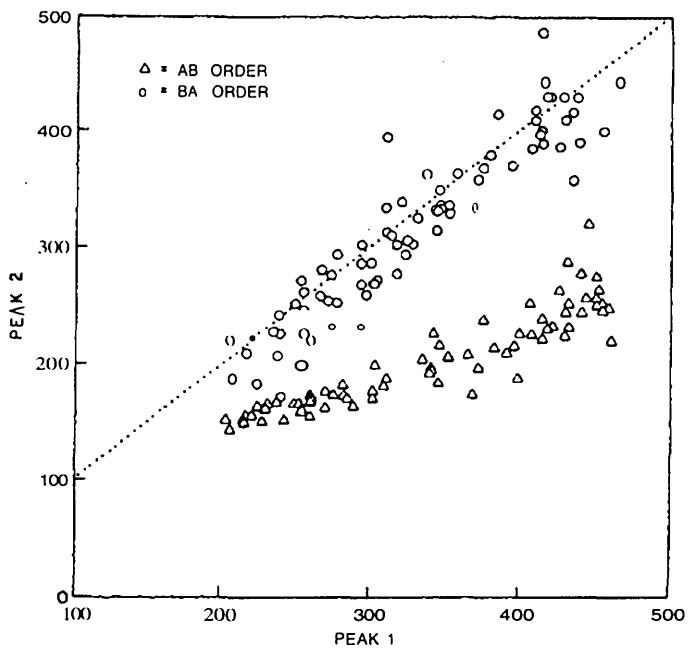


Figure 3.7: Data for one speaker in Pierrehumbert's (1980) *Anna/Manny* experiment. The $f0$ peak on *Anna* is plotted against that on *Manny* in backgound-answer (BA) and answer-background (AB) order for each rendition (Liberman & Pierrehumbert 1984, p.173).

3.1.4 Global Pitch Variation

The ToBI system assumes a perceptual distinction between *intrinsic* and *extrinsic* variations in pitch span (terminology as per Ladd 1996, pp. 269-83), i.e. variation relative to the pitch span of the current phrase versus variation in the pitch span itself.[8] It has been shown that when speakers raise or lower their voices the resultant scaling of tonal targets is remarkably consistent (Liberman & Pierrehumbert 1984, Rietveld & Gussenhoven 1985, Shriberg, Ladd, Terken & Stolcke 1996); as is the scaling of tonal targets across speakers (Ladd & Terken 1995). For example, Liberman & Pierrehumbert (1984) showed that, in cases such as (3.3), the peak on the *background* word was scaled consistently with that on the *answer* word, in either order, when speakers varied their overall emphasis on a ten point scale, see Figure 3.7.[9]

(3.3) a. *background-answer contour*

Q: What about Anna? Who did she come with?

A: (Anna) (came with Manny)

b. *answer-background contour*

Q: What about Manny? Who did he come with?

A: (Anna) (came with Manny)

Although this works well to describe pitch span variation between phrases, it assumes the pitch span stays constant over the course of each phrase. As pointed out by Ladd (1996, pp. 272-9), this leads to a rather anomalous treatment of downstep: i.e. the phonological scaling down of a peak relative to the preceding accent in the phrase in terms of variation in pitch accent *type* (e.g. see Figure 3.5). As we have seen downstep has consistent effects across all accent types (cf. Ladd 1996, ch. 3): it affects *phonetic*, but not *structural*, prominence (cf. Ayers 1996); and is claimed to mark relative givenness (Baumann 2005)

[8]Ladd (1996, pp. 260-1) also distinguishes *pitch span* and *pitch level*. The latter is broadly a speaker's reference point relative to which pitch span is determined. Covariation in span and level can be factored out fairly well using a logarithmic scale.

[9]Note that these authors claim that the peak height of the *background* word is scaled to the peak height of the *answer* word by the *same* factor in either order, plus a constant factor of *final lowering* which leads to the difference in the two orderings shown in Figure 3.7. In the next chapter we dispute this, and relate the difference to the effect of the place of the two accents in prosodic structure. However, it remains true that the relative scaling of the two accents is constant over different pitch ranges *within* orderings, which is the important point here.

Figure 3.8: Diagrammatic representation of intrinsic, and local and global extrinsic effects on pitch span, re Ladd (1996, ch. 7) (stars are strong nodes and parentheses phrase boundaries).

(see section 2.2.2.2).[10] Ladd (1996, p. 76) suggests it adds a nuance of finality or completeness. This treatment may explain why annotators find the distinction between downstepped and regular accents difficult. Silverman et al. (1992) found agreement rose from 61-67% to 73-79% if downstepped accents were grouped with their regular counterparts, e.g. H* with !H*. Pitrelli et al. (1994) and Syrdal & McGory (2000) report similar findings. Although this issue does not turn out to be critical for us, the treatment in Ladd (1996, ch.7) seems more consistent. He distinguishes between *local* and *global* extrinsic effects, see Figure 3.8. *Local extrinsic* effects describe the variation in pitch span within and between phrases, including downstep. *Global extrinsic* effects describe widening or narrowing of the pitch span; are gradient and directly signal paralinguistic 'meanings' such as speaker involvement and emotional state (see section 2.3.3).

A closely related issue is the treatment of *declination*. That is, the height of successive peaks in declarative utterances usually declines over the course of a phrase. This declination resets slightly at PhP boundaries, but can continue over larger phrase structures (see Figure 3.9). Unlike with downstep, the later accent is not perceived as lower; although declination may be meaningful, e.g. questions often have no declination, and it seems to be controlled to convey discourse structure (Sluijter & Terken 1993, Swerts 1997). However, it remains uncertain how it is best treated. For our purposes, we take it as a factor to control when trying to gauge the perception of pitch in declarative utterances.

[10] This is consistent with the noted ambiguity between heavy downstepping and deaccenting (Beckman 1996, pp. 46-51).

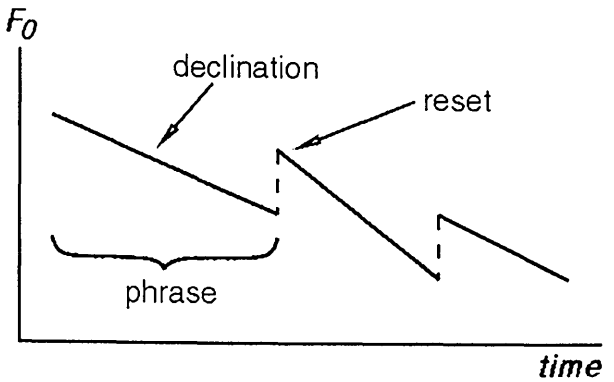


Figure 3.9: Pitch declination over multiple phrases (King 2001).

More important to our theory is the status of *emphatic accents*, or *expansions* in pitch span within a phrase (i.e. the opposite to downstep), e.g. *We were just sitting there, having a nice dinner, and then he BROKE UP with me*. We saw in the last chapter that these can signal a *restricted* kontrast interpretation. We develop this notion and discuss the status of emphatic accents at length in section 3.2.4. For now, we suggest they are consistent with a local extrinsic expansion in pitch span (as with downstep). These complementary relative height relationships between accents are important to signal information structure properties. Finally, it should also be noted that pitch is a key correlate to relative prominence within and across phrases, interacting in complex ways with the signalling of local and global extrinsic pitch span effects.

3.2 The Relationship Between Prosodic and Information Structure

In the last chapter, we set out the phenomena that we are trying to explain in this thesis, which broadly comprise the question of how information structure is signalled prosodically. We laid out the basic properties of this structure that need to be explained, i.e. focus/kontrast, focus projection, theme/rheme status and the role of contrastive accents. We saw that many of the uncertainties in the semantic account arise directly from (mis-)understandings of the prosodic facts to be explained; in particular, the nature of pitch accents. In this section, we argue that all of these information structure properties are signalled by relative prominence and phrasing within metrical prosodic structure. We show that many of the difficulties with

standard approaches to the signalling of information structure disappear when the full expressive power of this structure, as we have just set out, is taken into account. Importantly, we conceptualise this relationship as a probabilistic mapping between prosodic and information structure, because of the interacting influence of other factors on each (see further section 3.3). Finally, at the end of the last chapter we reviewed theories claiming that information structure is in part signalled by tonal event type. The implication of our theory is that intonational tune is much *less* important to signalling these meanings than is often claimed. Therefore, during our discussion we will try to suggest where the intuition behind these theories may have come from. We go into this question more deeply in section 3.3.

3.2.1 Association of Focus and Nuclear Prominence

In section 2.2.1, we defined focus in terms of the F-marking of elements in a clause, claiming that F-marking introduces a presupposition of an alternative set to the F-marked element, i.e. it marks kontrast (as per Rooth 1992). Elements which are not F-marked are interpreted as relatively given. As we saw, in standard accounts kontrast is marked by accenting, according to a *accent-first* theory of phrasal stress; along with *focus projection* rules which determine the scope of the focus using syntactic criteria. We showed that there are major difficulties with this approach, however (see also review in Ladd 1996, ch. 6). Focus projection rules have never been especially successful in explaining the patterns of obligatory and optional accents that occur in natural language. That is, as we saw in numerous examples throughout section 2.2, some accents seem to be either optional, or even obligatory, outside of focussed constituents. While most focal projection rules allow some "accents for rhythmical reasons", there is usually no explicit attempt to explain when and why these would occur, lessening the explanatory power of these theories. There are certain cases where focus seems to project from elements in a way clearly not allowed by standard theories. Finally, as we discussed in section 2.2.2.1, there are some foci, particularly 'given foci', which seem to occur without accenting, which cannot be explained by focus projection.

Many of these difficulties disappear if we take the relationship to be between F-marking and nuclear prominence, rather than with accenting per se. That is, assuming a *stress-first* theory of phrasal stress, we claim there is a strong constraint aligning kontrast with nuclear positions in metrical prosodic structure. As we shall see, when this view is taken, "accents for rhythmical reasons" are straight-forwardly accounted for by metrical structure itself. And focus projection rules turn out not to be necessary as focus scope is directly determined by the scope of nuclear prominence in phrasal structure.

Let us return to the basic facts that theories of F-marking and focus projection are trying to capture, using the example from section 2.2.1, repeated here as (3.4)-(3.8):

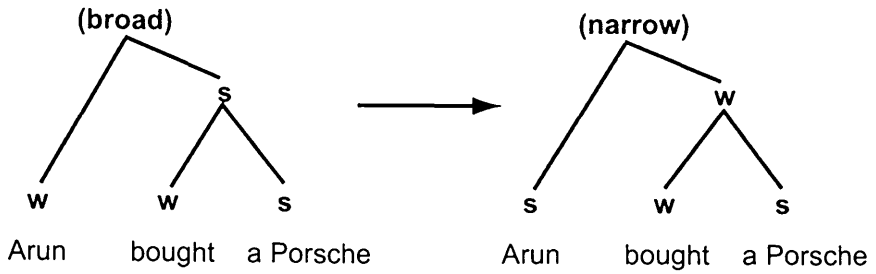


Figure 3.10: Reversal of relative metrical strength in the answers *ARUN bought a Porsche* and *Arun bought a PORSCHE* (adapted from Ladd 1996, p. 230).

(3.4) What did Arun buy?
(Arun bought a PORSCHE)

(3.5) What did Arun do?
(Arun bought a PORSCHE)

(3.6) What happened?
(Arun bought a PORSCHE)

(3.7) What happened?
*(ARUN bought a Porsche)

(3.8) Who bought a Porsche?
(ARUN bought a Porsche)

In the standard account this distribution of accents would be explained in terms of F-marking (shown by the accent) on the answer-word, and then syntactic projection from the object word to other parts of the clause in (3.5) and (3.6). Projection is not allowed from the subject onto the clause in transitive sentences, cf. (3.7). In *stress-first* accounts this is given a quite different explanation. The accenting pattern results from reversing the default *w-s* pattern of phrase level prominence to a *s-w* pattern, as we can see in Figure 3.10 (see Ladd 1996, ch. 6). That is, the relative prominence pattern in a phrase is constrained by the kontrast status and relative givenness of the elements of that phrase. In (3.4)-(3.6), the nuclear accent maps onto *Porsche* because it is kontrastive. *Porsche* has scope over the whole phrase, since it is the most structurally prominent, eliminating the need for syntactic focus projection rules. Further, a pre-nuclear accent on *Arun*, which sounds natural to many

speakers, would not be problematic for our theory, whereas it would for focus projection accounts such as Selkirk's (1995) or Gussenhoven's (1999*b*).

There have been several proposals along these lines in the literature recently (Ladd 1996, Truckenbrodt 1995, Truckenbrodt 1999, Büring to appear, Wagner 2006). For instance Büring's (to appear) proposal, which draws strongly on Truckenbrodt's (1995), claims that the highest ranked constraint on prominence is the following (in Optimality Theory terms):

(3.9) FOCUS PROMINENCE (FP) (Truckenbrodt 1995)
Focus needs to be maximally prominent.
A prosodic category C that contains a focussed constituent is the head of the smallest prosodic unit containing C.

Prominence is defined on prosodic structures. Büring assumes three levels of structure: the prosodic word, accentual domain, and the intonation phrase.[11] So, in (3.4), each accentual domain introduces a beat in the metrical grid for that utterance (taking the case where *Arun* is accented). In order to satisfy FP, the head of the PhP, i.e. the most prominent accentual phrase, must be the phrase containing the F-marked word, as we can see in (3.10) (adapted from Büring to appear, p. 17).

(3.10) Q: What did Arun buy?

	(					*		$)_{PhP}$
	(	*	$)_{AD}$	(		*		$)_{AD}$
A:	(Arun	$)_{Pwd}$	(bought	$)_{Pwd}$	([a Porsche	$]_F$		$)_{Pwd}$

Importantly, Büring recognises that kontrast is just one constraint on the realisation of prosodic structure, along with syntactic and general phonetic features. In the next section we take this idea further, showing how syntax, information status and phonetic factors interact in the realisation of prosodic structure; and how kontrast status is therefore interpreted from relative prominence patterns, rather than being directly signalled by them. In section 3.3 we argue that these interacting constraints should be modelled as probabilistic variables in the construction of metrical structure (in particular see Figure 3.15). Such an approach seems to scale up better as a whole language model than OT-constraints which are usually defined in a limited domain. Here, we continue our argument as to why prosodic structure is a better predictor of focus scope than syntactic projection rules.

[11] We do not include accentual phrases as there is little independent evidence for their existence in English, as noted above (Beckman & Pierrehumbert 1986). However, it seems compatible with this analysis to assume these accents are motivated by strong metrical nodes, without any prosodic boundary (cf. Beckman 1996, pp 38-41).

Büring (to appear) demonstrates that, adopting Truckenbrodt's (1995) proposal, focus can be *vertically projected* from any constituent, not just from internal arguments (as claimed by Selkirk); and that focus can be *horizontally projected* from any argument to its head, not just predicates (as claimed by Gussenhoven) (see also Büring submitted, Truckenbrodt 2006). For example, focus can project vertically from the subject in the following example (Büring to appear, p. 7):

(3.11) Q: Why did Helen buy bananas?

A: [(Because JOHN bought bananas)]$_{FOC}$

A': [(Because John is HUNGRY)]$_{FOC}$

Unlike the case with (3.8) above, (3.11A) is a focus on the whole clause, as the *why*-question presupposes a whole proposition response, not a focus on *John* (as shown by the comparison with (3.11A')). Projection from the subject position clearly violates Selkirk's (1995) rules. However, in our account it is straight-forwardly predicted by metrical reversal because of relative givenness. The nuclear prominence on *John* extends to the whole phrase, marking it as a Focus.

The metrical account also elegantly explains some of the more problematic distribution facts for focus projection rules. As we saw in sections 2.2.1.2 and 2.2.2.1, there is an asymmetry in the accentual marking of both predicates in 'all-new' sentences (e.g. (2.24) and (2.26) versus (2.28) and (2.30)) and given material (e.g. (2.49), repeated in (3.12)) in the pre- and post-nuclear region (Wagner 2006), i.e. in both cases these items are optionally accented before the F-marked element, but compulsorily deaccented after it:

(3.12) Arun bought a red Porsche. What did Joel buy?

a. (Joel bought a [GREEN]$_F$ porsche)

b. * (Joel bought [GREEN]$_F$ PORSCHE)

c. (Joel bought a green [MERCEDES]$_F$)

d. (Joel bought a GREEN [MERCEDES]$_F$)

As Wagner argues, if the F-marked element is mapped onto the nuclear accent, then this asymmetry has a unified explanation, following naturally from phrase level prominence structure, i.e. pre-nuclear material can be pitch accented, post-nuclear material cannot (as in Figure 3.4) (Wagner 2005, Wagner 2006). This asymmetry also explains the difference in realisation of early and late focus, i.e. to get an interpretation such as in (3.8), a large accent on *Arun* is needed, along with almost no pitch movement on *Porsche*; whereas an interpretation such as (3.4) can arise from a relatively smaller peak on *Porsche* and a moderate peak

on *Arun* (Rump & Collier 1996, Xu & Xu 2005). Further, as pointed out in Ladd (1996, pp. 228-31), the metrical account more easily explains the distribution of accents in certain syntactic structures. For example, while in most cases the accent is moved to the left (e.g. *Arun/Porsche*), in other cases it is shifted rightward (from Ladd 1996, p. 229):

(3.13) A: Where did you go just now?

B: (I took the GARBAGE out)

(3.14) A: What happened to all the garbage?

B: (I took the garbage OUT)

We would argue that this is because *garbage* and *out* are immediately dominated by the same node, so that metrical reversal leads to *out* and not *took* being accented. Focus projection theories are drawn into complicated explanations in terms of movement and trace marking to account for such cases.

3.2.2 Pre- and Post- Nuclear Accents and Ambiguity

In the above, we claimed that the appearance of pre- and post- nuclear accents is not problematic for our theory as they are expected as part of metrical structure. We need to refine this claim, to say that these accents are expected given the other known constraints on the appearance of strong nodes in this structure. As we saw in section 2.1, these include part-of-speech type, e.g. nouns are more likely to be prominent, verbs less likely; and syntactic constituency, e.g. heads are less likely to be prominent; as well as the general constraints on prominence structure laid out above, e.g. a right-branching bias, and rhythmic requirements (see further discussion of relevant constraints in Büring to appear). Therefore, for instance, the disputed evidence as to whether an accent is required on the subject in broad focus sentences (cf. Selkirk 1995, Gussenhoven 1999*b*), e.g. on *Arun* in (3.6) above, is directly related to whether the subject is "prosodically heavy", not to the meaning of the sentence. We would expect deaccenting on pronouns or other short noun phrases, but not on longer, more complex, subject phrases.

It follows from this that pre- and post- nuclear prominence *can* signal kontrast where this prominence is *not* expected. Where prominence patterns are not predictable, it makes the construction *marked*, which, as we argue further in section 3.3.1.1, leads to a kontrastive, or *restricted* kontrast, interpretation on non-nuclear elements. This is not incompatible with our claim above; we are merely saying that the focussed constituent is the head of a unit smaller than the PhP, as the nucleus is already occupied by another focus. For example, if

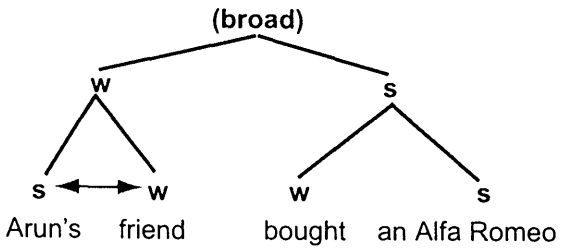


Figure 3.11: Reversal of relative metrical strength in the pre-nuclear domain in (3.15) to signal kontrast.

the pre-nuclear phrase is long, the distribution of accents may indicate kontrast, i.e. there may be metrical reversal in the pre-nuclear domain.

(3.15) What happened?
(ARUN's friend bought an ALFA ROMEO)

This is unambiguously a kontrast on *Arun*, i.e. as opposed to other people's friends, because the default stress would be on *friend* (see Figure 3.11). The appearance of stress on *Arun* is marked, leading to a kontrastive interpretation. Secondly, if a word which would normally not be prominent is accented in the pre-nuclear domain, this indicates additional meaning, such as kontrast (see further in section 3.3). So in the classic example, the fact that pronomial *he* is stressed accounts for the kontrastive meaning (*Bill* as opposed to *John*) (assuming *he insulted* is implied by *x called a Republican*):

(3.16) John called Bill a Republican, and then...
(HE insulted HIM)

The same holds in the post-nuclear case. In section 2.2.2.1, we reviewed a study by Beaver et al. (2004) showing post-nuclear given foci are marked by greater intensity and duration, but not pitch movement. As they suggest, this can be interpreted as association with the strongest point of metrical prominence in the post-nuclear domain (cf. Huss 1978). Ladd (1996, p. 227) gives a similar explanation to account for cases such as the following:

(3.17) A: Bill says you haven't helped him on his project very much.

B: I don't know what he's complaining about. I wrote an entire PROGRAM **for 'im**.

(3.18) A: Bill seems to think you've been giving priority to other people in the department.

B: I don't know what he's complaining about. I wrote an entire PROGRAM **f'r him**.

So, in (3.17), *him* is relatively given; and *for* kontrastive (i.e. opposed to *not doing things not for him*). Therefore *for* is associated with a strong node, and pronounced with a full vowel, while *him* is reduced. In (3.18), on the other hand, *him* is kontrasted (as opposed to all the other people in the department), and so it is said with a full vowel. These types of distinctions cause major problems for accent-first accounts, which cannot capture the marking of post-nuclear given foci.

However, taking this point to its logical conclusion, there is potential for ambiguity, at least in the pre-nuclear domain. For example, we saw in our discussion in section 2.2.1.2 that (2.21) (repeated below) is ambiguous between the reading in (3.19a), where *mother-in-law* is given, and that in (3.19b), where it is kontrastive (because of the parallel with *father-in-law*).

(3.19) What did Arun's mother-in-law think?

a. (Arun's MOTHER-in-law DISAPPROVED)

b. (Arun's MOTHER-in-law DISAPPROVED)
(but his FATHER-in-law LOVED it)

Here we claim that this ambiguity is part of the expected ambiguity of language, and may be disambiguated by context (as in (b)). More generally, this type of ambiguity is resolved by taking into account how likely it is that *mother-in-law*, said with a particular set of prosodic properties, is kontrastive, given its other semantic and phonetic properties. So here, the accent on *mother-in-law* in (3.19a) is unlikely to signal kontrast, even though it is given (cf. (3.16)), because the phrase is long. It could be disambiguated with a phrase break: (3.20) is much more likely to lead to the interpretation that *mother-in-law* is kontrastive than (3.19a).

(3.20) (Arun's MOTHER-in-law) (DISAPPROVED)

Note, however, that phrase length itself puts pressure on phrasing, so the utterance is not totally unambiguous. We will return to this in the next section.

The discussion in this section brings us back to some of the evidence presented in sections 2.2.1.2 and 2.2.2.2. That is, whether it is necessary to appeal to general syntactic and phonetic constraints to explain accentual patterns, or whether all of these factors can be boiled down to a relationship between prominence and informativity or predictability. We would argue that the evidence presented so far in this chapter shows that the constraints

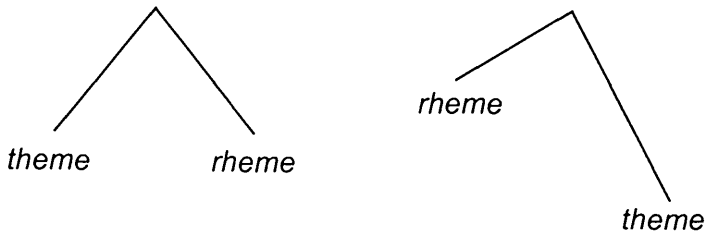


Figure 3.12: Diagrammatic representation of the signalling of the relative metrical prominence of theme and rheme nuclear accents.

on prosodic structure itself, and the effects these have on the interpretation of the scope of kontrast given prominence marking, exist apart from such notions. With regard to semantic constraints, in this work we largely assume that there are independent low-level syntactic constraints, such as the 'prominence-lending' character of object phrases over subjects, or nouns over verbs. These could explain differences in the prominence structure of, e.g. in intransitive sentences such as (2.23) and (2.29) in section 2.2.1.2. For instance, in *my CAR broke down*, the reversal of the usual weak-strong prominence pattern is not marked (and therefore does not imply narrow focus), because *car* is both an object (semantically) and a noun. On the other hand, *JESUS wept* would be marked because *Jesus* is the semantic subject. However, it is certainly plausible that informativity or predictability in general (apart from relative givenness) are further constraints on this structure. For the reasons given in sections 2.2.1.2 and 2.2.2.2, we do not believe these are the *only* constraints. However, as we discuss in Chapter 6, our theoretical framework, and the methodology presented there, can at least provide a means to assess the relative importance of these different factors in explaining relative prominence patterns.

Finally, there is also evidence that speakers use acoustic prominence independently in such cases to signal kontrast, especially where an element is not 'heavy' enough to form its own phrase (and therefore be associated with nuclear prominence). Studies have shown pre-nuclear accents in narrow focus sentences may be less prominent than in broad focus ones (Xu & Xu 2005, Jaeger & Wagner 2003). Further, Rump & Collier (1996) (discussed above), showed the 'optimum' realisation of 'double focus', i.e. two kontrasts in a phrase, was with both a high pre-nuclear and nuclear accent. An exaggerated accent on *mother* in either (3.19a) or (3.20) would increase the likelihood of a kontrastive interpretation.

3.2.3 Theme/Rheme Status by Relative Prominence

In the previous section, we passed over the status of these 'double foci' phrases. However, if we look at the relevant cases, we find that the kontrast in nuclear position is *rhematic*, and the less prominent kontrast *thematic*. So, for example in (3.17) and (3.18), *program* is in nuclear position because it is rhematic, providing new information in relation to the proposition of *helping Bill*. Similarly, Beaver et al. (2004) explicitly state that their analysis only applies to given foci. This brings us back to the issue raised in section 2.2.3, and that is the prosodic realisation of thematic and rhematic kontrasts, which are said to be distinguished by tonal accent type (e.g. L+H* versus H*) (Jackendoff 1972, Steedman 2000, Büring 2003). Here, we claim that this distinction is in fact signalled by structural relative prominence. As we have just seen, many themes are given and therefore do not form their own prosodic phrase. In these cases rhematic kontrast is signalled by association with nuclear prominence; although kontrast within the theme may be indicated by pre- and post-nuclear prominence patterns. A central claim in the present work is that this relationship also holds above the phrase level. That is, when the theme and the rheme each form their own phrase, the nuclear accent in the theme phrase is less structurally prominent than the nuclear accent in the rheme phrase. The signalling of this relationship mirrors that at the phrase level as described in section 3.1.2, see Figure 3.12. Returning to the example in (2.65), we see that theme/rheme status (θ/ρ) is reflected in the metrical structure, as follows:

(3.21) Moana and Geoff met at the train station

(						*	$)_{PhP}$
[(	*	$)_{PhP}$ $]_\theta$	[(			*	$)_{PhP}$ $]_\rho$
([Moana $]_F$	$)_{Pwd}$		(was going $)_{Pwd}$	(to [Paris $]_F$			$)_{Pwd}$

In this example, the accent on *Moana* would either be equally, or slightly less acoustically prominent than that on *Paris*. In order to signal rheme-theme order, the nuclear accent on the second phrase would have to be much less acoustically prominent than on the first, as in the following variation on (3.15):

(3.22) Who bought the Alfa Romeo then, if Joel and Arun both bought a Porsche?

If we look back at the contours in (3.3) above, we see that this is the relationship Liberman & Pierrehumbert (1984) found between peaks in their "background-answer" sentences.

We will see in the next chapter that, although this was not the aim of their study, their results in fact provide direct support for our position.

In order to substantiate why we believe thematic kontrast is signalled by relative prominence, not accent type, we need to separate out carefully what thematic kontrast is not. Normally, themes form part of the presupposition, and are not prominent. If they do appear in nuclear position, it is because their status is marked, i.e. particularly emphasised. Therefore, there is a strong correlation between thematic kontrast and emphasis, and the marking of emphasis needs to be carefully separated from the marking of themehood. As we set out in section 2.2.3, many descriptions conflate the marking of thematic kontrast and *restricted* kontrast under the term 'contrastive' accent. *Restricted* kontrast, re our definition, has also been argued to be marked by L+H*. However, as we will discuss more in the next section and show in the experimental work in the next chapter; once *restricted* kontrast is accounted for, the two turn out to be distinguished by structural prominence.

It would round off the argument to be able to explain where the intuition comes from that thematic kontrast is marked by accent type, particularly by a 'scooped' accent. We do not have definite answers to this yet, but we believe one reason speakers mark thematic kontrast in nuclear position is because of the expressive power of nuclear, as opposed to pre-nuclear, accents (see section 3.1.3.1). That is, it may have to do with illocutionary and affective connotations correlated with themehood, that are marked on the nuclear accent, rather than theme status per se. As we have just said, there is a strong correlation between the marking of kontrastive themes and emphasis, so the phonetic cues claimed to mark themehood may have as much to do with marking emphasis. Further, as we suggested in section 2.3.3, many utterances that have been claimed to be 'isolated themes' (particularly by Steedman), may in fact be instances of more general rhetorical subordination relationships between clauses, such as Nucleus-Evidence (cf. Mann & Thompson 1988). We will discuss this further in section 3.3, and explore it using examples in Chapter 7. On the other hand, we should note here that our claim is not necessarily incompatible with Steedman's claims about discourse semantics discussed in section 2.3. We disagree with him that the prosodic distinction between themes and rhemes is one of pitch accent type. However, it is still possible that his general scheme holds if this distinction is marked by relative prominence, but his other dichotomies (speaker/hearer supposition, and polarity in the common ground) are marked prosodically as he claims. Signalling of his 'isolated themes' would be more problematic, as in our scheme the theme prosodic marking is inherently *relational*; although it may be possible that speakers lower their pitch on such 'isolated themes' to show that they are prosodically subordinate to an unstated rheme. As we said in section 2.3.3, however, Steedman must still deal with the other contrary phonetic evidence for this to be the case.

To complete this argument we should note that the interpretation of kontrast status for themes in nuclear position is mediated, as with pre-nuclear accents above, by the likelihood that that theme would have formed its own phrase anyway. So, as we saw in (3.20), even with a phrase break the utterance is ambiguous as to whether there is a kontrast on *mother-in-law*; as the subject phrase is long and therefore there are independent reasons for a phrase break. Similarly, we do not expect all rhematic nuclear accents to be kontrastive either: if the rheme phrase is particularly long, or as we saw in (3.2), particularly emphatic, the kontrast in the rheme may also be signalled by nuclear prominence across multiple phrases. Again, this falls out from our notion of markedness. As we saw in section 2.2.3.1, information structure places a strong constraint on phrasing. However, this interacts with other inherent constraints, such as syntax structure, i.e. certain clause types are more likely to form their own phrase; and phrasing itself, such as physiological constraints on phrase length and eurhythmic effects.

Our theory allows us to suggest a different analysis for a number of cases which have recently been claimed in the literature to show evidence of 'nested foci', i.e. rather than a single dimension of focus, or two dimensions of theme/rheme and kontrast/background structure as assumed here, foci can be 'nested' within each other (Neeleman & Szendröi 2004, Féry & Samek-Lodovici 2006). We will see that this analysis also brings us back to the question raised in section 2.2.3.1 about the differing interpretations of relative givenness and kontrast. For instance, Féry & Samek-Lodovici (2006, p. 141-2) analyse the following response in terms of the nested focal structure shown, indicated by the metrical relationship (example from Neeleman & Szendröi 2004, p. 149):

(3.23) Father: What happened?

Mother: You know how I think our children should read decent books.
Well, when I came home, rather than doing his homework,

(						x						) IP
(	x		)	(		x	)	(		x		) PhP
[	Johnny	was		[	reading	Superman$_{F3}$		to some		kid	]$_{F2}$	]$_{F1}$

Féry & Samek-Lodovici (2006) claim that the whole clause forms a focus because of the question *What happened.* However, we would say that *Johnny* is clearly thematic, being set up as the topic by the preceding part of the *Mother's* reply. Therefore this utterance involves an information unit consisting of the theme *Johnny was* and the rheme *reading...kid.* There is a single 'nested focus' within the rheme, i.e. *Superman*, the kontrast. The rest of Féry & Samek-Lodovici's (2006) examples can likewise be analysed according to this two dimensional structure without the need for 'nested foci'. More interesting for us is the relative status of *Superman* and *some kid.* Féry & Samek-Lodovici (2006) claim this is a

single focus on *Superman*. However, we would suggest that with a particularly exaggerated accent on *kid*, the alternative set would include not only alternatives to *Superman*, but also to the person *Johnny* was reading to. That is, the salient properties of the alternative set within either the theme or the rheme are influenced not only by the position of the nuclear accent, but also by the relative prominence (structural and acoustic) of the other entities involved. Returning this to the question of the relationship between kontrast and relative givenness, we would say that increased prominence increases the likelihood of a kontrastive interpretation, as opposed to a relative givenness interpretation. Further, relative prominence within the theme/rheme unit can indicate not only the relative givenness of different elements, but also whether they form salient properties of the alternative set, e.g. whether the alternative set here is { *Superman*, *War and Peace*, *the Ascent of Man*, ... } or { *Superman to some kid*, *Superman to a Hollywood scout*, *War and Peace to his teacher*, ...}. We will see this more clearly using examples from our corpus in Chapter 7.

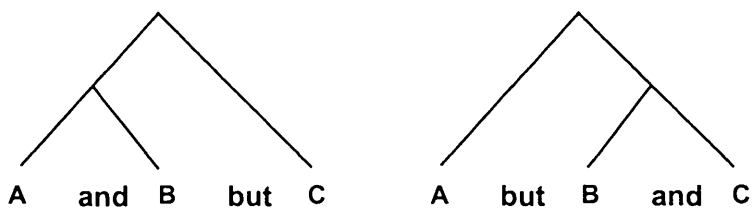
Finally, we can find independent evidence for the general claim that focus extends across phrases in the re-interpretation of two studies which aimed to show pitch scaling effects across phrase boundaries, but were not specifically looking at information structure. Ladd (1988) looked at the height of $f0$ peaks in sentences like the following (Ladd 1988, p.532):

(3.24) Allen is a stronger campaigner, **and** Ryan has more popular policies, **but** Warren has a lot more money. (*and/but* structure)

(3.25) Ryan has a lot more money **but** Warren is a stronger campaigner, **and** Allen has more popular policies. (*but/and* structure)

He hypothesised that the relative peak heights in each clause would reflect register differences in their relative attachment, i.e. since *and* is a closer connector than *but*, as follows:

(3.26)



As can be seen in Figure 3.13, results generally supported Ladd's hypothesis, i.e. pauses were longer between clauses connected by *but* than *and*, and the height of the initial peaks of the second and third clauses (B1 and C1) varied as expected. This was not true, however, of the subsequent peaks in the relevant clauses (B2, B3, C2, C3). These seemed to follow

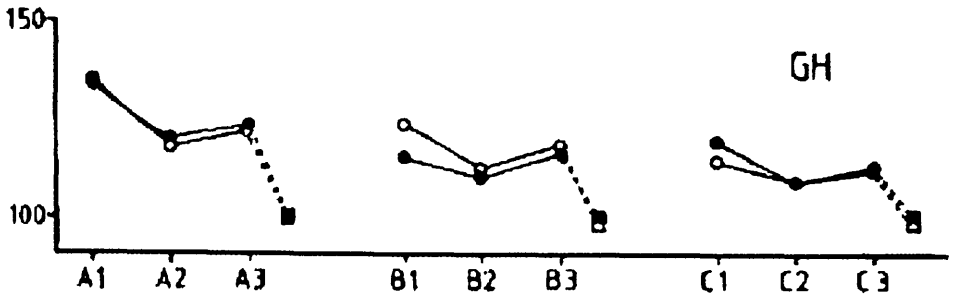


Figure 3.13: $f0$ peaks on accents (Hz) from one speaker in *and/but* clause structure experiment. Accents are numbered consecutively 1-3 within each clause A-C. Filled circles/squares show the *and/but* condition, hollow circles/squared the *but/and* condition (from Ladd 1988, p. 534).

a general declination pattern over the course of the utterance. Ladd claims this is because hierarchical relations between clauses are primarily signalled at the beginning of a phrase. However, the sentences naturally lend themselves to a contrastive interpretation between clauses, i.e. they were of the form *A is X but B is Y and C is Z*, leading to a contrast between Person A, Person B and Person C. The kontrast relationship of these entities could be captured by the scaling of these accents between clauses.

A similar argument can be made for Truckenbrodt's (2002) study of 'upstep' in some German dialects. In a production study, Truckenbrodt found speakers of these dialects raise the height of the nuclear pitch accents in non-final PhPs so they are comparable to the utterance-initial peak, against the downstep pattern in the rest of the utterance. For instance, in (3.27), *Leinen* is scaled equally with *Manu*, whilst the intervening accents are downstepped (Truckenbrodt 2002, p.93).

(3.27) Der MANU und die HANNE sollen der LENA im JANUAR das LEINEN weben,
und der WERNER soll in MURNAU MARONEN holen.
Manu and Hanne are supposed to weave the linen for Lena in January,
and Werner is supposed to get sweet chestnuts in Murnau.

Truckenbrodt takes this as evidence that the register of the last accent in the first PhP is associated with the register of the higher IP, rather than the downstepped from the preceding accent. He draws from this a generalisation that "pitch accents are phonetically scaled to the register that is correlated with the highest prosodic level they are associated with" (Truckenbrodt 2002, p.113). However, these effects did not hold across all speakers.

A possible reason for this is speakers' information structure interpretation of the utterances. Speakers were asked to read as if in response to the question *Was gibt's Neues* (*what's new?*), meant to ensure a broad focus reading. Again, however, the sentence form *A does X and B does Y* lends itself to an interpretation of a contrast between the X done by A and the Y done by B. That is, 'upstep' is the marking of kontrast across phrases. This analysis is supported by studies Truckenbrodt himself cites showing narrow focus blocks the application of downstep (Beckman & Pierrehumbert 1986, Féry 1993). The finding did not hold for all speakers suggesting not all had the same information structure interpretation. Of course, if the mapping between prosodic structure and information structure is probabilistic, as we claim; this result could also be explained in terms of expected variation in the prosodic realisation of a given information structure. However, since the speakers that did not produce 'upstep' varied systematically (following a general downstep pattern over the whole utterance), the result is more consistent with two competing underlying information structures.

3.2.4 Emphatic Accents and *Restricted* Kontrast

In section 3.1.4 we saw that pitch range can be raised (or widened) over whole phrases (c.f (3.1)). This raising is gradient, and directly linked to meaning, i.e. the wider the span, the more surprised, excited, etc. the speaker is. However, we saw that the range of a single word can also be raised, making it emphatic. In the literature there is evidence for a semi-categorical distinction between normal accents and these emphatic ones. We suggest the effect of these accents is to induce a *restricted* kontrast interpretation (see section 2.2.1.3).

In a series of experiments Ladd and colleagues showed that listeners may process accents differently depending on whether they perceive them to be *normal* or *emphatic* (Ladd 1993, Ladd et al. 1994, Ladd & Morton 1997). In Ladd et al. (1994), they replicated Gussenhoven & Rietveld's (1988) study showing that, in two peak utterances, lowering the $f0$ of the *first* peak (P_1) *decreased* the perceived prominence of the *second* peak (P_2). However, they found this effect only when P_2 was low. When P_2 was *high*, the effect was reversed: lowering P_1 *increased* the perceived prominence of P_2, see Figure 3.14. Ladd (1993) suggests that this contradictory effect is because listeners process P_2 as normal in the first case, and emphatic in the second. Therefore they perceive P_1 as having the same pitch span as P_2 in the first case, but not the second.

Ladd & Morton (1997) then tried to prove the existence of a categorical boundary between normal and emphatic accents using established tests (see Harnad 1987). Firstly, listeners heard utterances such as *the ALARM went off*, where the peak on *alarm* varied. Judgements on whether utterances were an "everyday occurrence" or an "unusual experience" were broadly S-shaped, consistent with a categorical boundary. However, in the second

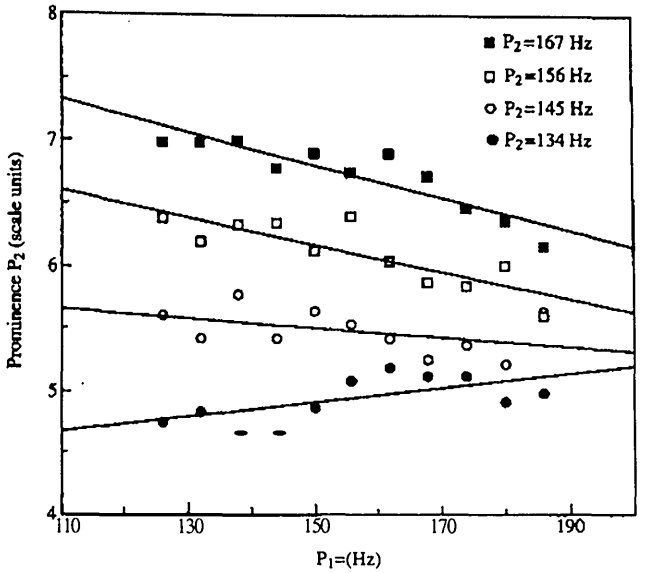


Figure 3.14: Ratings of the perceived prominence of P_2 as the $f0$ peak of P_1 increased, for different levels of P_2 (from Ladd et al. 1994, p. 97).

experiment, same/different judgement on pairs of utterances showed only weak evidence for a categorical discrimination peak. Ladd & Morton (1997) conclude that the distinction may not be categorically *perceived*, but may be "categorically *interpreted*". Actually, given our understanding of pitch range effects, these results are exactly what we would expect. Raising the peak in a single-accent phrase is ambiguous between raising the pitch span of the whole phrase and making that accent emphatic. In the first experiment, subjects had a context, i.e. "everyday" versus "unusual", which biased a normal versus emphatic interpretation and therefore led to a categorical perception boundary. In the second experiment, subjects had no context, and so not surprisingly used their ability to gradiently discriminate phrasal pitch range from peak height (e.g., as shown for P_1 in Ladd et al. 1994) confounding the categorical discrimination. At least in the case of prosody, the perception/interpretation distinction is not clear (see further in section 3.3). Pitch range is interpreted differently at different levels of structure; our primary cue to this level is the meaning it conveys (c.f Gussenhoven 1999*a*). Of course, this does not make it a simple problem, as an accent can both be both emphatic and raised for emphasis. However, these are potentially separable effects (cf. Rump & Collier 1996, Gussenhoven 1999*a*).

In fact, there is considerable evidence for a sharp interpretative boundary between normal and emphatic accents. So while a nuclear accent on an object is ambiguous between a narrow

and broad focus reading (e.g. (3.6) versus (3.4)), an emphatic accent on the object is unambiguously narrow focus (Eady, Cooper, Klouda, Mueller & Lotts 1986, Horne 1988, Rump & Collier 1996, Xu & Xu 2005). Rump & Collier (1996) showed 'optimal' realisation of narrow focus on the object was when the second peak was considerably higher than the first; as well as the most agreement on a narrow focus reading between different focus conditions (see further in section 4.1). As discussed in section 2.2.1.3, we can conceptualise this in terms of how restricted the alternative set is. Taking the *Arun/Porsche* example again, as an answer to *What did Arun do?* the alternative set could be *went to the gym, flew to Greece, ran for PM*; but for *What did Arun buy?* it is clearly more restricted (although both are theoretically infinite). We submit that emphatic accents disambiguate *restricted* and unrestricted kontrast readings. In the next chapter we see this analysis may help explain the disagreement as to the phonetic character of kontrastive theme accents: in some cases, they are compared to rheme accents that convey *restricted* kontrast, and in other cases they are not.

In section 2.2.1.3, we showed that *restricted* kontrast can lead to exhaustive or scalar implicature. The implicature has been claimed to be marked by L+H* (Pierrehumbert & Hirschberg 1990). This could because L+H* is being used to mark emphatic accents. The ToBI guidelines state L+H* is "a high peak target on the accented syllable... immediately preceded by relatively sharp rise", whereas H* is any rise "includ[ing] tones in the middle of the pitch range" (Beckman & Hirschberg 1999). In some formulations the distinction has been unambiguously defined in terms of peak height (e.g. Watson et al. 2004). That is, the nearest category to the semi-categorical distinction between normal and emphatic accents in ToBI is H* versus L+H*. Another source of confusion may be the substitutability of late peaks (also associated with L+H*) and high peaks. Late peaks have been linked to a *restricted* kontrast interpretation of themes in German (Braun 2005). Gussenhoven (2002) claims speakers can use, and listeners interpret, late peaks as being substitutable for high peaks, since in general high peaks are late, taking longer to reach. He gives examples of languages that mark narrow focus with late peaks. Once learned, this information can be used as a 'shortcut' in production. As we will see in the next chapter, this may be part of the story in the realisation of theme/rheme and kontrast.

This analysis is useful beyond the L+H*/H* distinction. For instance, Ward & Hirschberg (1986) found the L*+H LH% contour is associated with both *uncertainty* and *incredulity* (see section 2.3.3). In Hirschberg & Ward (1992), they found the strongest cue distinguishing the readings was peak height, with a low +H peak signalling *uncertainty*, and a high peak *incredulity*. They claim this variation is gradient consistent with degrees of 'speaker involvement'. However, as Ladd et al. (1994, p. 316) point out, *uncertainty* and *incredulity* readings seem pretty discontinuous:

(3.28) A: I hear John and Mary are calling it quits.

B: They're SEPARATING
L*+H LH%

With a normal peak the interpretation is "well, they're only separating, they may get back together" (*uncertainty*). However, with a high +H peak, there is an implication of surprise "do you really mean to tell me they're separating" (*incredulity*). The discontinuity might be better explained in terms of the scope of the focus, in the first reading it is the whole sentence, i.e. *they're separating* as opposed to *they're getting back together*, whereas in the second reading, there is narrow focus on *separating* (as opposed to *not separating*).

Finally, it should be noted that emphatic accents are not the only way to signal *restricted* kontrast. As we discussed above, a marked rendition of a word, i.e. more prominent than expected given its properties, makes such an interpretation is more likely. Further, it can come out of the context itself, without any special prosodic marking. For instance, in the *Arun/Porsche* example, if the general conversation were about people buying cars, the answer to *What did Arun do?* would probably have a restricted alternative set of {*bought a Porche*, *bought a Volvo*, ...}. However, an emphatic accent may still have the effect of forcing such an interpretation, or inducing a scalar interpretation.

3.2.5 Interacting Phonetic Cues

It is probably fair to say that the concentration in research on prosodic signals of meaning has been on pitch variation (see discussion in section 2.4). The vast majority of the experimental work, including most of that reported above, has only measured and manipulated $f0$ values. As we have seen, the prevalence of ToBI has further led to a large amount of attention being paid to the location of $f0$ target points. We end this section by noting that this concentration may have downplayed the importance of other phonetic cues in at least two key ways: firstly, the other correlates of our perception of prosodic prominence; secondly, other cues that signals the kinds of illocutionary and affective meanings attributed to ToBI tunes (see sections 2.3.1, 2.3.2 and 2.3.3).

In section 3.1.2, we argued for the *stress-first* theory of phrasal stress, i.e. pitch accenting is one cue to phrase-level prominence, rather than phrase-level prominence being marked by pitch accents per se. Although many researchers would probably agree that prominence is a complex amalgam of increased pitch, intensity and duration; most of the work reported above on the semantic importance of relative prominence has concentrated on differences in peak height (e.g. Ladd (1996, ch. 6), Rump & Collier (1996)). Impressionistically, in conversational speech some speakers (particularly male) are much more reliant on cues such as

lengthening and intensity to convey prosodic prominence, whereas others rely more on pitch movement (as we shall see in Chapter 7). State-of-the-art pitch accent detection algorithms report improved performance when using combinations of $f0$, intensity and duration features than when using pitch alone (Conkie et al. 1999, Chen & Hasegawa-Johnson 2004). Further, emphatic accents have been found to be predicted by mean intensity, $f0$ excursion and syllable duration, in that order (Brenier, Cer & Jurafsky 2005). Lengthening has been shown to be a good cue to prominence in perception studies (Carlson & Granström 1986, Aylett & Turk 2004). Mean intensity has been found to be greater in accented than unaccented syllables (Sluijter & van Heuven 1996). In a recent study of a wide variety of dialects of English, Kochanski et al. (2005) found that the 'prominence' of syllables (as marked by both trained and naive annotators) is best indicated by an amalgam of loudness and durational cues, with loudness being more important. $f0$ level and movement were relatively unimportant in their models.[12] As mentioned above, in the absence of $f0$ cues, listeners can detect nuclear accents on the basis of lengthening and intensity cues (Hermes & Rump 1994, Turk & Sawusch 1996). We have seen that prominence distinctions signalled only by intensity, duration and vowel quality can be meaningful in the post-nuclear domain (see section 3.2.2). Lastly, recall that in section 2.2.2.2 we showed that these features may be partially independent, with duration being more closely linked to predictability, and pitch and intensity to informativity (Watson & Arnold 2005).

In the last chapter we reviewed claims that illocutionary and affective 'meanings' are conveyed by intonational tunes, or combinations of tonal events. These accounts have intuitive appeal, but they are based on a perceptual understanding of the realisation of phonological intonation events. The phonetic properties of these realisations relevant to signalling these 'meanings' probably extend beyond the definitional cue of tonal target alignment. No doubt intonational tune is part of the story. However, there has been surprisingly little work on whether other phonetic cues, particularly lengthening, intensity and voice quality, are as important, if not more important, to convey the intended connotations.[13] There are some indications, however, that they could be. We saw in section 2.3.2 that in the automatic recognition of dialogue acts, $f0$ cues were not more effective, and in many cases were less effective, than other prosodic cues. Overall, duration features were used in 50% of decisions, while $f0$ features in only about 10% (Shriberg et al. 1998). Hirschberg & Ward (1992) found that voice quality was a significant cue to the distinction between the *uncertainty* and *incredulity* readings of L*+H LH% (see last section). Intuitively, this makes sense: a tremulous voice can

[12] Although it should be noted that these authors made a binary distinction between 'prominent' and 'not prominent' syllables. It may be that $f0$ is more important to signalling *levels* of prominence.

[13] A notable exception to this is the studies by Scherer et al. (1984) and Ladd et al. (1985), reported in section 2.3.3.

convey uncertainty even in a straight declarative statement. However, Hirschberg & Ward (1992) did not test whether listeners perceived a statement to be uncertain in the *absence* of $f0$ cues.

3.3 The nature of Prosodic Units

In the foregoing discussion, we have described how the AM framework models prosody in terms of a metrical structure of prominences and phrases, and a sequence of tonal events associated with prominent nodes in this structure, the intonational tune. In the last section, we claimed that the mapping between metrical structure and the segmental string is constrained by information structure, as well as lexical and syntactic distinctions. Most of the research cited above assumes, without much argument, that there are a limited number of rules or principles (possibly in terms of OT constraints) by which this mapping can be determined. Stemming from work in the computational field, this type of assumption has been seriously questioned in many areas of linguistics in recent years. Below we will set out evidence for the probabilistic processing of language in general, both in computational applications and as reflecting cognitive reality. We will use this work to argue that the realisation of prominence and phrasing structure should also be modelled probabilistically.

We have seen that intonational tune is claimed to signal illocutionary and affective meanings, and (disputedly) information structural distinctions (e.g. kontrastive theme accents). However, as we discussed, there is still no general agreement as to whether these meanings are derived compositionally from the basic meanings of intonation events, from the intonation contour as a whole or somewhere in between. Here we will argue that this confusion may stem from assumptions about both compositionality and categorical perception which have been challenged in recent years. We will argue that the meaning of a contour is neither compositional nor holistic, because of the high mutual information of each of its parts. We will then move on to look at the events themselves. Recent work has questioned traditional wisdom about both categorical phonemic perception and phonemic categories in general, showing that people store remarkable levels of detail about instances of a category. We will discuss the implications of this for our concept of intonation categories, given their ideophonic nature (see section 2.3.3), suggesting that potentially all deviations in the production of a certain category are meaningful.

3.3.1 A Constraint-Based Approach to Information Structure Interpretation

It is fair to say that stochastic approaches have become ubiquitous in the fields of computational speech and language processing (e.g. see discussion in Jurafsky & Martin 2000, ch. 1). With recent increases in computer power, the ability to model the effects of large numbers of variables over huge amounts of data has meant that these models usually easily outperform rule-based approaches. We have seen prosodic events are predicted using stochastic methods from a combination of lexical and acoustic features in most speech recognition and synthesis applications. In general, linguistics has been slow to accept that this evidence might reflect the stochastic nature of human cognitive processing, emanating from Chomsky's (1957) attack on such approaches. He argued that since language is creative and infinite, it cannot be probabilistic; but based on inherent grammatical rules, as in his famous example *colorless green ideas sleep furiously*. However, this has come under increasing criticism with the success of computational stochastic language models. Grammaticality has itself been found to be a gradient concept directly related to frequency (Bard, Robertson & Sorace 1996, Sorace & Keller 2005). Furthermore, when multiple levels and types of constraints are taken into account, probabilistic models much more closely predict human behaviour than explanations based on the interaction of inherent rules as we shall see.

What are the implications of probabilistic cognitive processing? Language comprehension is the process of determining the most likely interpretation, rather than *the* interpretation, of a speech signal given all the information available, including the acoustic signal, the speaker, the setting and the preceding conversation; using our knowledge of the likelihood of each interpretation given this information drawn from our previous language experience. Processing difficulty and disambiguation should be directly related to frequency on every level, i.e. from recognition of phones to semantic and pragmatic interpretation. On the production side, the way a message is conveyed is strongly affected by the likelihood of the different structures and forms that are capable of expressing it; with planning difficulties associated with unlikely structures and forms (see review in Jurafsky 2003).[14]

There is much evidence to support this approach in language comprehension. Take the well-studied phenomenon of *garden path* sentences, e.g. *The horse raced past the barn fell.* That is, after the word *barn*, people show a processing difficulty because of the ambiguity between a main clause and reduced relative clause reading of *raced past.* Firstly, main clause constructions are more frequent than relative clauses. Further, effect does not hold if the verb more frequently occurs in a reduced relative clause (MacDonald & Seidenberg

[14] It could be argued that the formulation of messages themselves is affected by frequency in each speaker's experience, however this philosophical argument belongs in a different work.

1994, Trueswell 1996). It is also affected by whether the noun phrase is a 'thematic fit' for a particular reading, i.e. an animate subject biases a main clause reading while an inanimate subject biases a reduced relative clause (Trueswell, Tanenhaus & Garnsey 1994). It can also be overridden by prosodic information (Schafer, Speer & Warren 2000); and presumably by a disambiguating preceding context. In other words, multiple sources of information are being used to make decisions about the interpretation of the utterance, and that information is about frequency distributions. There has been less work done on production, since it is hard to simulate knowing a speaker's message, but not how they are going to say it. However, as noted above, there is a strong relationship between lexical frequency and reduction, controlling for multiple factors (Bell et al. 2003, Bell et al. 2004); and, as we saw in section 2.2.2.2, between accessibility and referring expressions, accenting and reduction. Bates & Devescovi (1989) showed the frequency of relative clauses in a language is related to their usage after controlling for semantic and pragmatic factors.

The effect of frequency from multiple sources of information can be described using *constraint-based* models (for a review of other models see Jurafsky 2003).[15] The idea is that probabilistic constraints are used to assess parallel competing interpretations (e.g. Tabor, Cornell & Tanenhaus 1997). Experimental evidence comes from regression analyses showing the effects of these constraints, which we employ in Chapter 6. In production, it is argued that frequency increases activation, and therefore the likelihood of words and structures used (Roland & Jurafsky 2001). The importance given to a constraint is related to its *validity* (Bates & MacWhinney 1989). Validity is computed as a combination of the amount of the time is it available, how reliable it is and how often it also a cue to a different interpretation.

Bringing this back to the interpretation of information structure, we have shown that this structure is subject to multiple constraints, which have the capacity to be modelled probabilistically. We have seen that hierarchical prominence and phrasing structure strongly constrain the interpretation of information structure. However, the interpretation of this structure is also affected by the likelihood of the prominence and phrasing properties of the words in it. For example, function words are less likely to carry stress than content words; and prominence varies by part-of-speech type. Syntax strongly constrains phrasing, although with some clause types more than others. Interpretation is also evidently affected by context, i.e. the likelihood of any information structure interpretation given the current discourse model. Topic structure is also indicated by prominence and phrasing, which in turn affects the reliability of these cues to information structure. Further, prominence and phrasing structure may be manipulated along with melody to convey illocutionary and affective connotations related

[15] We should note that these models provide a way of thinking about and testing ideas about probabilistic language processes, they do not claim to actually describe cognitive processing (see note in Jurafsky 2003, pp. 89-90).

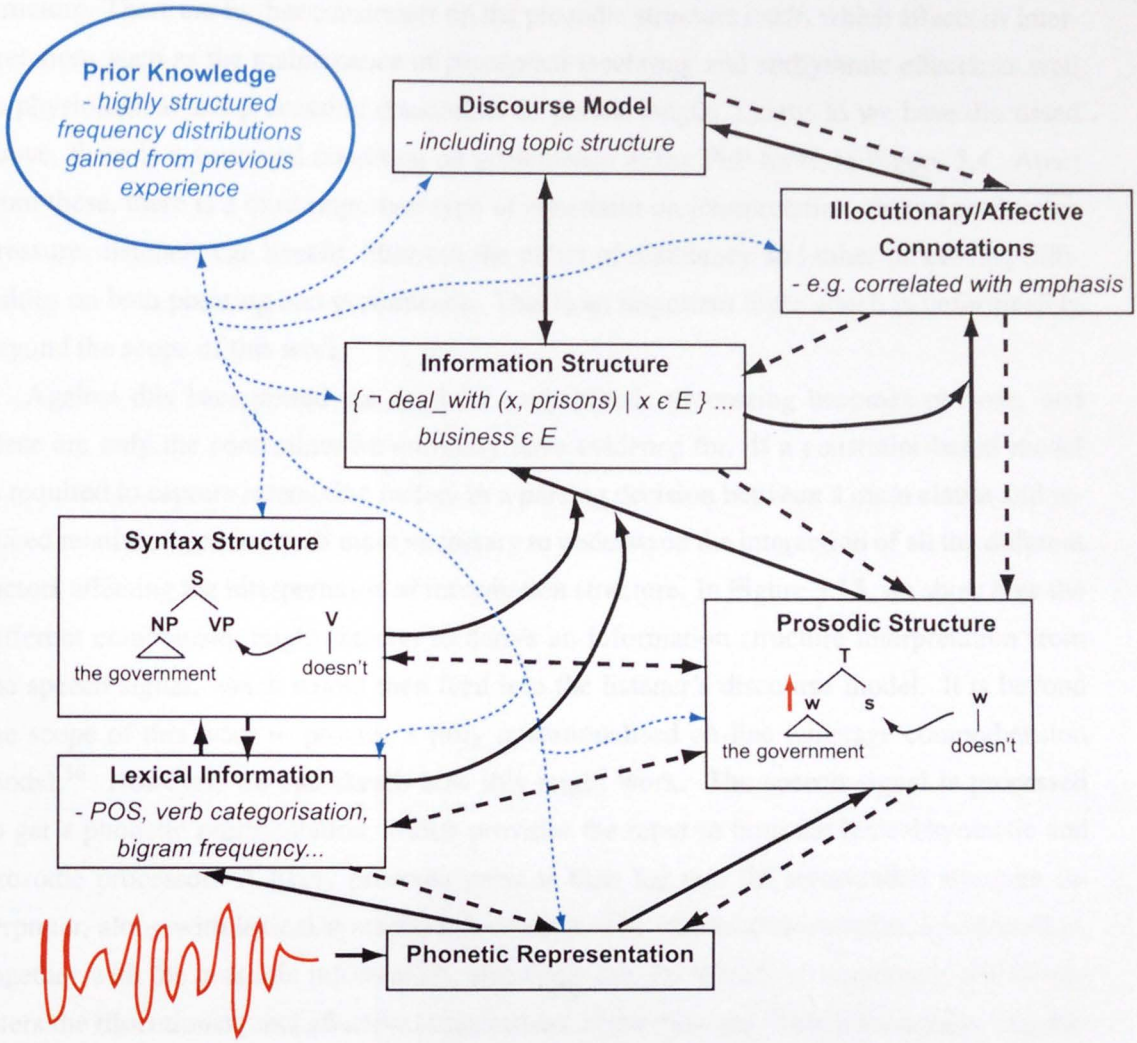


Figure 3.15: Diagram showing constraints on the parsing of prosodic structure, and the interpretation of information structure (intonational tune is not shown). Note that Information and Syntax Structure could be merged if we assume a syntactic parser such as CCG where the two are isomorphic (see Steedman 2000). A solid line shows that that component acts as a strong constraint on the parsing/interpretation of the component it points at, while a dashed line indicates a weaker constraint. The bi-directional arrows between each component and Prior Knowledge are meant to show that the interpretation in each component is determined by the probability of different interpretations given the frequency distributions for different inputs, and the constraints inherent on that component (e.g. rhythmic effects); as well as that each new utterance affects future prior knowledge. The utterance is *the government doesn't have to deal with it*, which is also used in Figure 5.1 and discussed in section 7.3.

to emphasis, which, as we have discussed, may or may not be directly related to information structure. There are further constraints on the prosodic structure itself, which affects its interpretation: such as the maintenance of perceptual isochrony and eurhythmic effects; as well as physiological and processing constraints on phrase length. Lastly, as we have discussed above, there is a structural condition on prominence at the PhP level, re Figure 3.4. Apart from these, there is a third important type of constraint on interpretation, that of production pressure: listeners can usually filter out the effect of disfluency and other processing difficulties on both phrasing and prominence. This is an important topic which is unfortunately beyond the scope of this work.

Against this background the need for probabilistic processing becomes obvious, and these are only the constraints we currently have evidence for. If a constraint-based model is required to capture interacting factors in a parsing decision between a main clause and reduced relative, it seems much more necessary to understand the interaction of all the different factors affecting the interpretation of information structure. In Figure 3.15, we show how the different components might interact to derive an information structure interpretation from the speech signal, which would then feed into the listener's discourse model. It is beyond the scope of this work to provide a fully operationalised on-line language comprehension model.[16] However, we can sketch how this might work. The speech signal is processed to get a phonetic representation, which provides the input to both the lexical/syntactic and prosodic processor. A likely prosodic parse is then fed into the information structure interpreter, along with lexical/syntactic information. The information structure interpretation, together with the prosodic information, also feeds into an 'affective' interpreter, which registers the illocutionary and affective connotations of the message. This information, together with the information structure, is then used to update the discourse model. As a probabilistic computational model, we envisage that each component computes multiple parses (or representations) at once, maybe even passing these simultaneously to the higher components. The plausibility of each parse is continually measured from further information coming upward from the advancing speech signal, but also downward as feedback from other components in the system (the dashed lines). So for instance, ambiguity in prosodic phrasing might be resolved by the syntactic phrasing, if this was more certain; and the information structure interpretation is also strongly affected by plausibility in the context. Within each component, the likelihood of each parse/representation is computed from the input given the known constraints on interpretation for that component. These may be inherent, e.g. eurythmic effects on prosody, or caused by interactions between components, e.g. parsing preferences for different verbs.

[16] For example, this could also be envisaged as a Bayesian belief network (see Jurafsky 2003).

Figure 3.15 shows a model of the comprehension of information structure. As we noted, constraint-based models of language production are less developed. However, we would imagine that production of prosodic structure, given a certain information structure, would largely be the reverse of the structure shown. Prosodic structure would be realised incrementally with input from the information structure and affective components. It would be combined with the segmental string in the phonetic component. The only change would be that some sort of higher-level control would be needed to ensure the nuclear accent for the whole prosodic structure was in the right place for each information unit. Figure 3.15 also allows us to see how prosodic parsing itself is affected by interacting probabilistic constraints. That is, it is largely derived from the phonetic input, given the known structural constraints on prosodic structure itself. However, the probability of different prosodic parses is also affected by their plausibility given feedback from the lexical, syntactic, information structure and affective components.

3.3.1.1 Markedness

We have argued that *restricted* kontrast and kontrast within theme are marked concepts. Here we can clarify this statement. There is a long noted asymmetry between many pairs of related linguistic concepts, i.e. one is more *marked* than the other (e.g. see Jakobson 1963, Jakobson & Pomorska 1990). Thought to be a property of Universal Grammar, contradictory facts about similar oppositions in different languages were always problematic. Recently, however, it has been shown that markedness falls out easily from the relative frequencies of the pairs. That is, the unmarked member is more frequent, and therefore more predictable and more prone to reduction effects such as assimilation, neutralisation and underspecification. The marked member, on the other hand, is less predictable, and therefore needs to be more perceptually salient (e.g. see Hume 2004, Haspelmath 2006).[17] *Restricted* kontrast is marked in relation to 'ordinary' kontrast because, in the normal case, it is less predictable; likewise with non-kontrastive and kontrastive themes. We saw in section 2.2.1.3 that, theoretically, kontrast always involves the presupposition of alternative sets. However, in the case of rhematic kontrast, it is often not necessary to resolve the identity of this set in any detail. Therefore utterances with *restricted* kontrast readings are less predictable because this resolution is necessary. Themes are usually highly accessible in the context, therefore, if they are kontrastive, they are less predictable. This nicely supports our theory. A *restricted* kontrast or thematic kontrast reading comes about when an item is more prominent than ex-

[17]Note that Haspelmath (2006) argues that the term markedness should be dispensed with, because it can largely be reduced to frequency. However, we believe it is useful here as a way to think about the effect of increased prominence on the interpretation of kontrast, as long as it is understood to be a frequency-based phenomenon.

pected given its properties. Further, 'plain' rhematic kontrasts in nuclear position are more subject to other pressures on the prominence of items, e.g. downstepping due to accessibility (cf. Baumann 2005). That is, the constraint on the relationship between *restricted* kontrast and prominence is stronger than with kontrast in general. In Figure 3.15, this is shown by a red 'strengthening' arrow next to the weak node dominating *government* in the Prosodic Structure. This 'strengthening' leads to a *restricted* kontrast reading, which is shown by the salience of the alternative set { *business* } in the Information Structure (see further in section 7.3). As we will see in the next chapter, distributional evidence about the accentual marking of the two types of kontrast support this contention that *restricted* kontrast is a marked form of kontrast (e.g. Hedberg & Sosa 2001, Watson et al. 2004).

3.3.2 Intonational Tunes and Categorical Perception

In the discussion at the end of the last chapter, we saw that, despite such a long history of research, there is still no general agreement on how intonational tunes convey the types of illocutionary and affective meanings which they intuitively do. The main topic of this thesis is the prosodic signalling of information structure; and as should be clear from this chapter, our argument is that tune is much *less* important than some have claimed in its conveyance. However, our explanation would not be complete unless we at least sketch where the intuition that tune signals information structure comes from. As we will argue below, much of the difficulty comes from trying to frame intonation categories in terms of now disputed ideas about both morphemic and phonemic categories. Once a broader view of these categories is taken, an explanation of the certainties and subtleties of intonational tune seems closer at hand.

As we discussed in sections 2.3.2 and 2.3.3, there are two general approaches to how meaning is derived from intonational tunes (whether or not one assumes an underlying string of tones). The first is that the whole contour has a meaning (e.g. Fujisaki 1996, Liberman 1975). The second approach is to take each (ToBI) tonal event as morphemic, so the meaning of a whole contour is derived compositionally from the meaning of each event (e.g. Pierrehumbert & Hirschberg 1990, Steedman 2000). The whole contour approach has been criticised as being too brittle, i.e. unable to deal with other pressures on where accents are placed and meaning similarities between similar tunes (e.g. see Ladd 1983). As we have seen above, the tonal approach has as yet been unable to provide convincing evidence of the meanings of each of these intonational morphemes.

However, the idea that these approaches are mutually exclusive only holds if one assumes that meaning is, in most cases, strictly compositional. In the case of ordinary morphemes, it was standardly held that the meaning of a sentence is derived from the meaning of each of its

words, apart from a few 'idiomatic' phrases evaluated as a whole, e.g. *It's raining cats and dogs*. In recent years, however, it has been shown that 'idiomatic' usages are a much more central to language than previously thought. Jackendoff (1995) suggests they may double the size of the mental lexicon. Further, compositionality can be measured using the *mutual information* of word sequences, i.e. if words are more likely to occur together than either of them occur separately, then their meaning together is less likely to be compositional (e.g. Church & Hanks 1989). In fact, some words are very hard to define apart from their contexts, such as phrasal verbs involving *get*, e.g. *get up*, *get off*, *get over*, *get around*, etc. Recent cognitive models of the lexicon involve 'chunks' of various length stored in the brain and retrieved during production and comprehension, rather than a full derivation every time (see review in Sprenger, Levelt & Kempen 2006).

Now, it turns out that the mutual information value of ToBI events is generally high. Dainora (2002) found that the nuclear pitch accent is a very good predictor of boundary tone type in the Boston Radio News corpus (Ostendorf, Price & Shattuck-Hufnagel 1994). For example, 83% of boundaries following L* are H%, while only 17% after L*+H are. 39% of boundary tones following H* are high, compared to 54% for L+H*. The likelihood of an H* LL% tune is 33%. Dainora (2001) takes this to argue against the compositional approach. However, given our analogy with morphemic meaning, this is not necessary. Some tunes, e.g. H* LL% and L*+H LL% are sufficiently cohesive to be interpreted as a whole; whereas others, such as the continuation from an L+H*, are likely to be compositional. No figures were given for accent to accent probabilities. However, previous studies have shown that usually either all accents in a phrase are of one type, or all pre-nuclear accents are the same type and the nuclear accent is different (Crystal 1969, Ladd 1996, Dilley 2005). The change at the nuclear accent may help draw attention to it. The important point is that in the more predictable cases, like with *get on*, *get over*, etc., there may be some general sense in which the meaning is related to the component parts, however, we do not expect this to be actively used or even clearly statable.

The phonetic status of these intonation events is also problematic. If semantically they are analogous to morphemes, phonetically, they are analogous to phonemes, and have been partially subject to the same categorical perception tests (Pierrehumbert & Steele 1989, Ladd & Morton 1997, Redi 2003, Dilley 2005). These tests assume that our *perception* of speech sounds is biased by phonemic boundaries. That is, we cannot distinguish changes in a continuous phonetic variable (in the classic case VOT) unless it crosses a phonemic boundary. However, recent research has questioned this (e.g. Massaro 1998, Pierrehumbert, Beckman & Ladd 2000). Firstly, the tests do not work very well with vowels (Gerrits & Schouten 1998). More crucially, recent experiments have shown that performance in standard tasks is depen-

dent on internal subjective criteria, i.e. knowledge of phonemes (Schouten, Gerrits & van Hessen 2003). When a less biased task is used, e.g. comparing two sounds to a third, subjects can discriminate within-category stimuli. In other words, all categorical perception is in fact the interpretation of a continuous signal (see Gerrits 2001). This actually is not surprising: given the noise and ambiguity in the speech signal, it would seem odd that we would throw information away. This type of intuition lies behind exemplar theory, which has been used to explain certain facts about phoneme perception (see review in Pierrehumbert 2000). Far from all the instances of a phonemic category being equal, fine phonetic differences between different renditions of a category are stored in a multi-dimensional probability distribution for its later recognition; and this recognition is crucially dependent on context.

What are the consequences of this for our notion of an intonation event? Firstly, as we discussed in sections 3.1.3.2 and 3.2.5, theoretically, ToBI categories are described in terms of the association between syllables and tonal targets. There is no general agreement on what other cues listeners use to identify them, making recognition much less robust, cf. in recognising a /t/, listeners can use VOT, formant transitions, aspiration, etc., as well as the predictability of a /t/ given the context. Secondly, all variation in the realisation of an event is potentially meaningful at the same level of interpretation, unlike with phonemes. With a /t/, a longer than usual VOT might be interpreted as "/t/ at the start of a prosodic phrase", i.e., the contextual effect on realisation is at a different level of interpretation. However, variation in accent shape may affect the same level of interpretation, i.e. discourse semantics. Further, it may be directly meaningful. As argued by Liberman (1975), intonation is ideophonic, i.e. the relationship between the sound and the symbol is not arbitrary, because of the inherent 'meaning' of different pitch manipulations (cf. Ohala 1994, Gussenhoven 2002). As discussed in section 2.3.3, the events themselves may be grammaticalised, but these variations are not. If, for example, we take L* HH% as the grammaticalised contour associated with a *declarative question*, an especially low L* might be interpreted as something like "declarative question along with doubt or contradiction", i.e. effects on the *same* level of interpretation.[18]

This discussion does not imply, as Taylor (2000) claims, that tonal event shape should be described entirely by continuous variables as in his Tilt system. Some aspects of intonation event interpretation are clearly grammaticalised (see section 2.3.3). Further, recent work on the precision of tonal alignment suggests categorical use (see section 3.1.3.2). In fact, a phasing relationship between $f0$ and articulatory movement would tie in nicely with evidence showing languages exploit non-linearities in the physical system to form phonemic categories (see discussion in Pierrehumbert et al. 2000). Our suggestion is actually

[18] If we accept the distinction between illocutionary and affective connotations is hard to draw, and therefore that these operate on the 'same' level.

related to the 'categories plus features' proposals of Ladd (1983) and Gussenhoven (1984) (see section 2.3.1). However, there is no reason to presume these features are categorisable themselves, rather the 'features' could be gradient manipulations of tonal events.

The argument made here is that we may be able to keep a conception of intonation categories as being semantically similar to morphemes, and phonetically similar to phonemes; if we take on board all the implications of this, given recent developments in our understanding of the nature of both morphemes and phonemes. That is, intonational meaning is partly compositional and partly holistic. Intonation events should be defined by clusters of phonetic features; and variation in the realisation of these events may be meaningful at the same level of linguistic interpretation.

3.4 Summary and the Next Steps

In this chapter we have set out, drawing from evidence in the literature, our assumptions about the nature of prominence, phrasing and intonational tune. Using these concepts, we advanced our theory about the relationship between prosody and information structure. Information structure is a strong constraint on the mapping of the segmental string onto metrical prosodic structure. In particular, kontrastive elements try to align with nuclear accents. However, if the nuclear position is filled (with another kontrast), then kontrast may be signalled by relative prominence patterns in the pre- and post- nuclear domain, or by increased acoustic prominence. Structures with ambiguous interpretations may be resolved by phrasing or context. Theme/rheme structure has a strong impact on phrasing structure, so where there is a kontrast in the theme it wants to form its own phrase. Thematic kontrast is then distinguished from rhematic kontrast by relative prominence at the level of phrasing that contains the whole information unit. That is, themes are less structurally prominent than rhemes, rather than being marked with different pitch accent or boundary tone types, as was previously claimed. *Restricted* kontrast is a marked version of kontrast, and therefore is more likely to be realised with enhanced prominence, particularly with emphatic accents. Further, referent accessibility may have the effect of reducing acoustic prominence. Lastly, we looked at where the intuition that information structure is signalled by tonal events might come from, and suggested it may arise from an amalgam of certain information structure configurations and more subtle variations of different phonetic cues at the phrase level.

Along with this argument, we have tried to show why information structure needs to be conceived as a probabilistic constraint on the mapping of words onto prosodic structure, rather than determining it. Relative prominence, constrained by rhythmic requirements, can signal lexical stress distinctions, word class, syntactic attachment, accessibility, kontrast

status, rhematic versus thematic status and *restricted* kontrast, as well as be manipulated gradiently to signal general emphasis and different affective and emotive states. Phrasing, constrained by production and eurhythmic pressures, can be used to signal syntactic structure and information structure (theme versus rheme); as well as manipulated to signal emphasis and related affective connotations. $f0$ peak alignment, constrained by articulatory pressures, can vary depending on segmental make-up, word length, surrounding prosodic context, information status, illocutionary or affective connotation. As we have seen, most of the literature has concentrated on properties of the $f0$ contour; however, there is evidence that duration, intensity and spectral features are also used to signal both prominence and more subtle connotations.

In the rest of the thesis, we will test the predictions of our theory using quite different methodologies. In the next chapter, we report a number of experiments which directly test whether theme and rheme accents are distinguished by tonal accent type, i.e. L+H* versus H*, or by relative prominence. This question is central to showing why our *stress-first* approach explains the relevant phenomena better than an *accent-first* approach combined with a compositional semantics of tonal events.

Phonetic production and perception experiments are good to test particular distinctions predicted by different frameworks (Pierrehumbert et al. 2000). However, especially in work looking at high-level semantic properties, there is always uncertainty about how applicable results are to naturally occurring speech. Most importantly for our claims, such experiments are highly constrained in the number of variables which can reliably be tested at the same time, where it is precisely the impact of multiple constraints on prosodic realisation that we are interested in. In Chapter 6, therefore, we go on to use regression-based analysis of a small portion of the Switchboard corpus which we have annotated for relevant semantic and prosodic features. This analysis allows us to test whether broad distributions of semantic and prosodic features in our corpus are consistent with the predictions of our theory. We will see that there are limitations in this method as well: natural language data is messy and the production of annotated data time-consuming and necessarily inexact. It can also be hard to get at fine distinctions in our theory using this 'broadbrush' approach. Therefore, in Chapter 7, we conclude with a close analysis of examples from one conversation in the corpus. These examples can show strikingly the effectiveness of our theory over previous explanations of the marking of information structure; though of course such analysis cannot claim to have broad coverage of the data.

The use of a variety of methodologies is deliberate, as the short-comings of each approach can be compensated for by the strengths of each of the others, and corroborating evidence from different sources leads to greater confidence in the underlying theory being tested. It

is also probably an inevitable consequence of trying to test in a more broad-ranging manner claims about the realisation of phenomena, namely information structural properties, that are largely described in theoretical work drawing on introspective evidence (unlike the bulk of prosody research). While semantic theories are inherently difficult to test, in this study we particularly wanted to see if our claims would hold over a wide range of language data, i.e. spontaneous speech. The importance of making claims which are both verified and have broad coverage has been echoed in a number of recent work, though the articulation of their efforts may be quite different (Taylor 2000, Pierrehumbert et al. 2000, Dilley 2005, Xu 2005).

Chapter 4

Searching for Contrastive Accents

In the last chapter we described the basic elements of prosodic variation: prominence, phrasing and tune. Our contention was that the importance of prominence and phrasing in conveying meaning has been downplayed in comparison to the importance of tune, in part abetted by the widespread use of the ToBI annotation system. In particular, it has been claimed that intonational tune is important to signalling various aspects of information structure, as set out in section 2.3.1 (e.g. Pierrehumbert & Hirschberg 1990, Steedman 2000). Central to these sorts of proposals is the claim that *contrastive* elements are marked by a different type of pitch accent (or boundary tone) to elements conveying *new* information. As we discussed in section 2.2.3, *contrast* in fact covers two distinct discourse semantic notions: whether an element is Contrastive, evoking a *restricted* kontrast interpretation; and the marking of thematic kontrast.[1] It turns out that both of these notions have been claimed to be marked by L+H* (LH%), as opposed to the non-contrastive, rhematic H* (LL%). Unfortunately, as we saw in section 3.1.3.2, the distinction between L+H* and H* is one of the most contentious in the ToBI system, with the phonetic basis of the difference between them disputed and annotator agreement figures low.

In the first section below, we review previous experimental work relating to the realisation and interpretation of 'contrastive' accents and the L+H*/H* debate. We begin by setting out phonetic studies looking at the supposed basis of the categorical distinction between L+H* and H*. We then review various experiments looking at the production and perception of 'contrastive' versus 'informational' accents, either on the basis that these are L+H* and H* and/or using acoustic analysis. These studies suggest that the principal distinction is that contrastive accents are *higher* than informational accents. We then describe a series of production and perception experiments that we carried out looking at the realisation of thematic

[1] In this chapter, we capitalise Contrastive to refer to an element which explicitly contrasts with an equivalent element in the surrounding context, to separate this from kontrast, i.e. evoking an alternative set. Note all Contrastive elements are kontrasts, but not the other way around.

and rhematic Contrastive accents. A central conclusion from the first two experiments is that thematic accents are *lower* than rhematic accents. This finding was developed in the later two experiments to show persuasive evidence that theme/rheme status is indicated by relative prominence.

These experiments help locate information structure, the main focus of this thesis, as being primarily conveyed by variations in prominence and phrasing, not intonational tune. This finding is important to establish what prosodic variation is in fact due to intonational tune, and therefore what variations in tune are meaningful. This should help to place intonation coding systems like ToBI in their proper place within the overall scheme of prosodic variation.

4.1 Previous Experimental Work on Contrastive Focus

We have seen in the discussion over the past two chapters (particularly in sections 2.2.1.3, 2.2.3, and 3.2.4) that it is frequently claimed that 'contrastive' focus is marked with a different type of pitch accent to 'ordinary' focus, often identified as L+H* and H* respectively. However, these accents have very low annotator agreement, and the phonetic basis for the distinction is disputed (see section 3.1.3.2). Below we set out studies showing the supposed phonetic distinction between them in the ToBI framework is untenable, i.e. the temporal alignment and height of the L target. In the second section, we review conflicting evidence in the literature for an interpretative difference between the two, and conclude this difference must have another basis. Finally, we go through studies looking directly at the acoustic realisation of 'contrastive' versus 'ordinary' accents. During this discussion, we refer to our distinction between the marking of Contrastiveness and thematic kontrast. As we will see, this distinction is not often recognised in the literature, with many studies simply assuming one or the other represents 'contrastive' focus. We will try to identify which conception is used in each study in order to help us draw conclusions about the realisation of each which will be important to the design and analysis of our own experiments below.

4.1.1 Phonetic Characteristics of L+H* and H*

Theoretically, the distinction between L+H* and H* is that the former has a definite L target at the beginning of the rise to the H* peak, while the latter does not (see section 3.1.3.2). The problem with this is that if there is an accent which follows another rising accent, or the first accent is a few words into a phrase, be it H* or L+H*, the $f0$ curve tends to 'dip' before the accentual rise. This apparent low can be very hard to distinguish from an L target, e.g. see Figure 4.1. In her original analysis, Pierrehumbert (1980) analysed this 'dip' as a "sagging

Figure 4.1: Similarity of the realisation of the L target in an L+H* accent (on *Lombard*) and the 'sagging transition' before an H* accent (on *Lambert*) in one production taken from Experiment 1. $f0$ trace is the blue line and intensity curve the dashed red line, vertical lines show word boundaries and phone boundaries in the target words.

transition" between two H* accents, implying that it varies as a function of the number of unstressed syllables between the accents, rather than being a fixed target. However, even she recognised that this was problematic (Pierrehumbert 1980, p. 70).

In a series of production and perception experiments, Ladd & Schepman (2003) showed the existence of a fixed L target in H* accents, that is the start of the rise is reliably anchored with the beginning of the stressed syllable, regardless of the number of intervening syllables. Their first production study looked at the alignment of tonal targets in name pairs such as *Norma Nelson/Norman Elson*, produced with an H* accent on both names. The number of syllables in between peaks was varied using different names. The location of the L target relative to the beginning of the stressed vowel (V1) in the surname was consistently later in the *Norman Elson* cases than the *Norma Nelson* cases, regardless of the number of intervening syllables. A complementary perception experiment showed L location was a small, but significant, aid to name pair discrimination. A third study showed the scaling of the L target was not affected by the number of intervening syllables as long as there was at least one. Ladd & Schepman (2003) concluded that, since both accents involve an L target, they should be merged, although there may be an interpretative distinction in terms of gradient differences in peak height or valley depth.

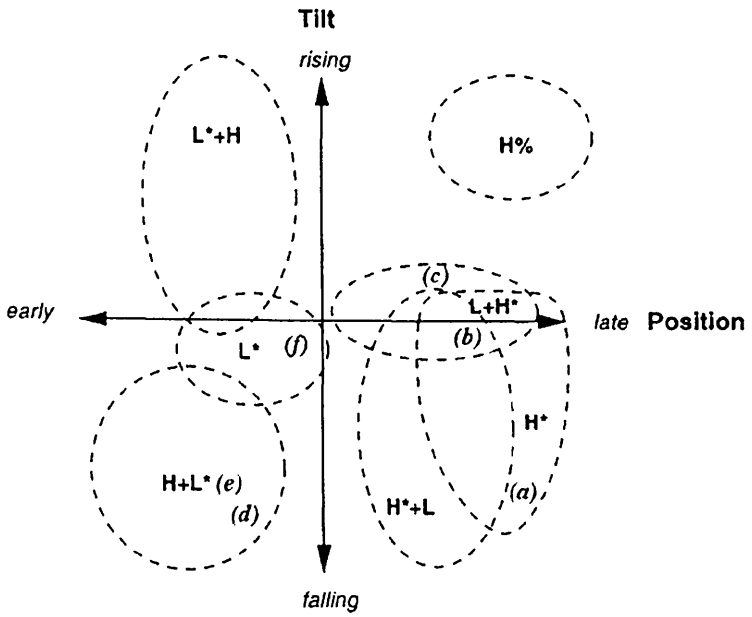


Figure 4.2: Two dimensional schematic representation of ToBI accents in intonational space in terms of the parameters of the Tilt system, i.e.: *rising/falling*, the proportion of the accent that is rising/falling; and *position*, the position of the peak. This shows the substantial overlap in the realisations of L+H* and H* (from Taylor 2000, p. 1711).

The ToBI guidelines define L+H* as "a relatively sharp rise from a valley in the lowest part of the speaker's pitch range", while H* is any other rise (Beckman & Hirschberg 1999). Dilley (2005) investigated whether there was therefore any evidence for a categorical boundary in the scaling of the L target. In an imitation experiment, subjects heard the phrase *some oregano* with the peak on *reg* held constant while the height of the preceding L was varied (*o'regano* in American English). Imitations varied gradiently, along with the stimulus, i.e. there was no interpretative boundary between a low and intermediate L target (cf. Pierrehumbert & Steele 1989, Ladd & Morton 1997, Redi 2003). Dilley (2005) concluded that her result supported Ladd & Schepman's (2003) contention that L+H* and H* should be merged. Finally, Taylor (2000) showed that there is substantial overlap in the phonetic characteristics of L+H* and H* accents as marked in the Boston Radio News corpus: that is, in the alignment of the peak relative to the stressed syllable, and the proportion of the accent where $f0$ was either rising or falling, see Figure 4.2.

We can see that it is hard to draw a categorical distinction between L+H* and H* on the basis of the alignment or scaling of the L target. Both accents have this target, and it has

been suggested that, on phonetic grounds, they should be merged. However, this still needs to account for the alleged interpretative distinction between the two. This, as was suggested in section 3.2.4, may in fact come from the height and/or alignment of the H* peak.

4.1.2 Interpretative Differences between L+H* and H*

We have seen that L+H* is linked to 'contrastive' readings, and H* to 'new information'. A number of studies have therefore looked for interpretative differences between the two accents. Hedberg & Sosa (2001) analysed the marking of information structure categories with ToBI accent types in a corpus of televised political talk shows. They investigated three dimensions of information status: ratified and unratified, i.e. mentioned before; topic and focus, equivalent to theme and rheme; and Contrastive. These formed five categories (it was assumed foci were unratified), a sample of which were ToBI annotated, see Table 4.1. Hedberg & Sosa (2001) conclude that there are correlations with information structure categories, but these are not straightforward. Most of the time all categories (except for ratified topics, which are usually unaccented) are marked with H*. They infer that L+H* is used for "emphatically highlighting an element relative to its context" (Hedberg & Sosa 2001, p.14), which covered their contrastive topic, unratified topic and contrastive focus (83% of L+H* tokens). L*+H usually marks topics, H* focus, and 'upstep' (<H) (a higher H* than the preceding one in that phrase) is also only found on foci. In our terms, we could interpret this as showing L+H* (and possibly 'upstep') mark *restricted* kontrast, rather than thematic kontrast. Consistent with our contention that this is a marked category, this is not obligatory, as most accents are H*. Unfortunately, however, since Hedberg & Sosa do not provide an acoustic analysis, it is not possible to ascertain the basis of their L+H*/H* distinction. However, we will see below that L+H* seems to be used elsewhere to mark increased relative prominence, so this finding is consistent with our claims. Further, in our scheme Hedberg & Sosa's (2001) association of 'upstep' with foci (our rhemes), could be analysed as marking the prominence of a Contrastive rheme in relation to a Contrastive theme, i.e. preserving the weak-strong relationship when both are emphasised. We find support for this idea in our results below.

Ito, Speer & Beckman's (2004) study looked at accent type in noun/adjective pairs. Subjects gave unscripted instructions involving different colours and types of ornaments. This allowed the adjective and the noun to vary between being *discourse given/new* or *Contrastive*, e.g. *blue bell/orange bell* or *blue bell/ blue house*. Theme/ rheme status was not reported on, and is difficult to establish in this context. Over 80% of both adjective and noun tokens were accented, except in *new-given* contexts (*blue bell/orange bell*) (less than 50% of nouns), and *given-given* (67% of nouns). Once more, we see the asymmetry in the marking of given

	H*	L+H*	L*	L*+H	ϕ
Ratified Topic	11	1	4	0	26
Contrastive Topic	24	10	1	2	5
Unratified Topic	23	13	0	3	2
Contrastive Focus	23	11	7	0	1
Plain Focus	27	6	8	0	0
Total	108	41	20	5	34

Table 4.1: Distribution of pitch accents (or their absence) relative to information structure type (modified from Hedberg & Sosa 2001, p.12).

items in pre- and post-nuclear position (see section 3.2.2). They further report that L+H* is more likely to be used in Contrasts than H*. However, its distribution is also asymmetric. Around 50% of Contrastive adjectives are L+H*, while only around 20% of Contrastive nouns are (compared to 4-5% of each in non-contrastive contexts). A follow-up eye-tracking study also showed this discrepancy: while L+H* on Contrastive adjectives helped identify the correct referent; L+H* on Contrastive nouns did not significantly improve identification times (Ito & Speer 2005). Again, they do not give an acoustic analysis of their material, leaving it open to re-interpretation, as we shall see in our discussion of Krahmer & Swerts (2001) below.

A number of studies, however, investigated the interpretation of L+H*, while giving some indication of its acoustic realisation. Bartels & Kingston (1994) looked at the importance of different supposed cues to the distinction between L+H* and H* in producing Contrastive and non-Contrastive versions of a referent. In the first experiment, subjects heard a conversation such as:

(4.1) Q: So, did Amanda eat anything today?

A: Yes, she ate her apple.

B: (AMANDA had a BANANA)

The peak height, L scaling, and the temporal alignment of the rise onset and peak on *banana* were systematically manipulated and the utterance resynthesised. Subjects judged whether B meant she had a *banana* rather than an *apple* (contrastive), or as well as one (non-contrastive). In the second experiment, subjects heard the same response and were asked to judge whether the question had been *What's Amanda been up to today?* (non-Contrastive)

or *So, did Amanda have any fruit today?* (Contrastive). Bartels & Kingston (1994) found no significant difference between the two experiments. For both, by far the most reliable cue was peak height, with Contrastive accents being higher, despite the fact this was varied only moderately (16Hz). L scaling and peak alignment both provided weak cues, used more by some subjects than others. The L on Contrastive accents was lower; and, surprisingly, the peak on Contrastive accents was earlier. This finding is further support for our contention in section 3.2.4 that *restricted* kontrast is marked by higher accents; as we would claim in both scenarios *banana* is a *restricted* rhematic kontrast.

In an eye-tracking study, Watson et al. (2004) measured subjects' fixations on objects in a visual display while hearing instructions such as:

(4.2) a. Click on the camel and the dog.

b. Move the dog to the right of the square.

c. Now, move the *camel/candle* below the triangle.

In the context, *camel* is Contrastive, and *candle* is not. The accent on *camel/candle* was either L+H* or H*. Watson et al. (2004) found that when the target had an L+H*, subjects at first fixated on Contrastive pictures, whether or not this was the actual referent. They only resolved the correct referent late in the production of the target word. When the target had an H*, subjects could resolve the correct referent quickly, with no preference for new pictures. Watson et al. (2004) concluded that the interpretative domains of the two overlap: L+H* creates a bias for Contrastive information, while H* is compatible with both new and Contrastive information. In our scheme, *camel* is both thematic and Contrastive, while *candle* is more ambiguous. The authors explicitly state, following Bartels & Kingston (1994), that their L+H* accents were realised with higher peaks than their H* accents. They did not investigate whether the Contrastive preference increases gradiently with peak height, or if there is an interpretative boundary correlated with peak scaling.

This phonetic detail may help explain the contrary result in Welby (2003). Welby investigated the acceptability of L+H* and H* in VP and (O)bject-focus sentences, as in the following:

(4.3) Q-VP: How do you keep up with the news?

Q-O: Do you read the Lantern?

A: I read the Dispatch.

She hypothesised that an L+H* on *Dispatch* would be dispreferred in the VP-focus reading (response to Q-VP), but there was no significant difference in appropriateness ratings for

either reading over H* renditions. However, Welby notes that she chose L+H* and H* tokens with equivalent peak heights so this would not form the basis for subjects' judgements. Given our discussion thus far, her contrary result result may show that it does.

From these studies we see that there is reasonable evidence Contrastive referents are more likely to be marked by L+H* than non-Contrastive referents. We take this as showing L+H* is being used to mark *restricted* kontrast. However, this marking is not obligatory, as in most cases Contrastive referents are marked with H*; and H* does not create an interpretative bias. From the available evidence the primary, if not the only, phonetic cue to the distinction is peak height, with L+H* accents being *higher* than H* accents.

4.1.3 Interpretation and Realisation of 'Contrastive' Accents

A number of other studies have looked directly at the realisation of 'contrastive' focus, without presupposing ToBI marking such as L+H* and H*. These studies help us to interpret some of the findings above, and flesh out the distinction between *theme* and *rheme* accents, and the marking of *restricted* kontrast. In an earlier study with a similar design to Ito et al.'s (2004), Krahmer & Swerts (2001) looked at the distribution and realisation of accents in adjective-noun pairs in Dutch. Subjects gave unscripted instructions about different coloured and shaped cards, e.g. *blue square*, *red circle*, again allowing adjectives and nouns to be defined as *new* (N), *given* (G) or *Contrastive* (C) (i.e. contrasting with the previous dialogue turn). Accenting was judged and reported by two experts separately. Their findings on accent distribution broadly accord with Ito et al.'s (2004). In the NN condition both adjective and noun were almost always accented (93/100%), and in most cases in the CC condition (56/69%). In the CG condition, the adjective was always accented and the noun deaccented; whereas in the GC condition, the noun was always accented while the adjective was also sometimes accented (31%), reflecting the usual asymmetry.

Krahmer & Swerts (2001) attribute differences in the shape of the accent on the adjective in the NN and CG condition not to accent type, but to structural accent position. In the NN condition the accent on the adjective is pre-nuclear, as the nuclear accent is on the noun; whereas in the CG condition the noun is deaccented, so the nuclear accent is on the adjective. This would concur with our analysis in the last chapter. Krahmer & Swerts (2001) go on to show that these Contrastive accents are not inherently higher or differently shaped to 'new information' accents, but rather contrastiveness is judged on the relative prominence of accents in an utterance. In a perception experiment, subjects were presented with pairs of utterances and asked to judge in which the noun or the adjective sounded most prominent. Single Contrastive accents (Cg and gC) were judged the most prominent (capitalised letters show the judged word); given items the least (Gc and cG); and CC and NN conditions intermediate.

However, when the relevant words were presented in isolation, prominence ratings changed dramatically. Both words in the NN conditions were judged most prominent, while the CC conditions were rated nearly as low as the given cases (Gc and cG). The single contrast cases were in between (Cg and gC). Acoustic analyses showed these isolated prominence ratings are generally correlated with the *f*0 height and intensity of the accents, showing the perception of prominence in their earlier experiment was *relative*. The distinction between the NN and CC conditions could give further support to the idea that discourse givenness is marked by reduced overall *f*0 levels (cf. section 2.2.2.2). Returning to Ito et al.'s (2004) study for English, if we accept L+H* is used to indicate that a referent is especially prominent relative to the context, the results in that study are consistent with Krahmer & Swerts's (2001) explanation. Further, this would neatly account for the discrepancy in L+H* marking on the adjective and noun, which Ito et al. (2004) could not explain.

Rump & Collier (1996) looked at the perception of focus scope in Dutch given the heights of the peaks on *Amanda* and *Malta* in (4.8) as a response to the following questions (see also section 3.2.4):

(4.4) What is happening? *(neutral focus)*

(4.5) Is John going to Cyprus? No... *(double focus)*

(4.6) Is John going to Malta? No... *(Sbj focus)*

(4.7) Is Amanda going to Cyprus? No *(Obj focus)*

(4.8) (AMANDA is going to MALTA)

Results showed that subject focus is signalled by a high peak on the first accent, and a much lower or absent second peak; whereas late single focus can be signalled by a high second peak and a lower, but moderate, early peak. Neutral and double focus are signalled by relatively equal peaks, though the second is slightly lower. The height of the first peak is the main indication of neutral (lower) versus double (higher) focus. The results for object and subject focus reflect the usual asymmetry. The result for double focus again shows a *restricted* kontrast interpretation *can* be distinguished from a 'neutral' interpretation by increased prominence. The design does not allow us to assess cues to the distinction between rhematic and thematic kontrast, however, as there is not enough context to judge theme/rheme status.

Braun (2005) looked directly at the realisation of themes and rhemes in Contrastive and non-Contrastive contexts in German. She analysed a distinctive tonal contour marking 'contrast' in German, the hat pattern. This is not very common in English, though her results

are still pertinent here. Subjects read short paragraphs where the target word, e.g. *the Malaysians*, was in either a non-Contrastive or Contrastive context, as in:

(4.9) Many Europeans don't know much about Malaysia. The country consists of two islands...
The *Malaysians* live from agriculture... (*non-contrastive*)

(4.10) Malaysia and Indonesia are neighbouring countries in the South China Sea... In Indonesia, tourism is very important and many people work in this sector.
The *Malaysians* live from agriculture... (*contrastive*)

Braun found that in both contexts, themes, e.g. *Malaysians*, were marked with rising accents and rhemes, e.g. *agriculture*, with falling accents. Contrastive themes were marked with higher and/or later peaks than non-contrastive themes; subjects seemed to have a 'strategy' for which marking they preferred. The accented syllable was also longer. Contrastive themes were usually followed by lower rheme peaks, while rheme peaks in non-Contrastive contexts varied. In a complementary perception experiment, listeners preferred utterances with high peaks in Contrastive contexts, although late peaks acted as a weaker, secondary cue; duration had no significant effect. This accords nicely with the suggestion in section 3.2.4 that *restricted* kontrast can be marked with high or late peaks. In the non-Contrastive context, however, all manipulations were equally preferred. Braun hypothesised that listeners either accommodated a Contrastive reading, or associated the high peak with greater speaker involvement, etc. This was confirmed in a follow-up perception experiment where listeners explained their choices. In non-Contrastive contexts, subjects chose the non-Contrastive contour (early, low peak) on information structural grounds, but the Contrastive contour because the speaker sounded more friendly, interested, etc. However, in the Contrastive context, they usually preferred the Contrastive contour on information structural grounds. This supports our contention in section 3.3 that *restricted* kontrast is a marked category; plain kontrast is more prone to variation because of other communicative functions. Braun also found that if the rheme accent was low, listeners were more likely to perceive the theme accent as Contrastive, showing again that pitch cues are evaluated relative to context. Finally, a number of trained GToBI annotators (Grice, Baumann & Benzmüller 2005) marked Contrastive and non-Contrastive theme examples. Braun found that, for the most part, the same label was used for both versions, and there was no general agreement on what this label was; further showing the difficulties with the ToBI system for capturing these meaningful distinctions.

Finally, Liberman & Pierrehumbert (1984) looked at the realisation of accents in contexts such as the following (see also section 3.1.4):

(4.11) Q1: What about Manny? Who came with him?

Q2: What about Anna? Who did she come with?

A: (ANNA) (came with MANNY)

Liberman & Pierrehumbert (1984) were interested in what they saw as universal properties of intonation contours, not discourse semantics. Therefore they saw this context as merely useful to elicit two accents of systematically different heights, which they analysed as both being H*. As an answer to Q1, *Anna* would be the *Background*, and lower than *Manny*, the *Answer* (BA order). As an answer to Q2, the roles, and therefore the relative heights, of the accents would be reversed (AB order). Under our analysis, *Background* is *theme* and *Answer* is *rheme*. Subjects were instructed to put a clear accent on both *Anna* and *Manny* in each case. That is, they were 'trained' not to produce a weak accent on *Anna* in BA order, nor to deaccent *Manny* in AB order; though the authors acknowledged this would also sound acceptable. We would suggest this forced a *restricted* kontrast reading for both *Anna* and *Manny* (cf. discussion in section 3.3.2). Liberman & Pierrehumbert (1984) found that, in both cases, the accent on the Answer was higher than the accent on the Background. However the relationship was asymmetric. In BA order the Answer accent was only slightly higher, in AB order it was much higher. Liberman & Pierrehumbert analyse this as the combined effect of the relative prominence of the two accents and 'final lowering', a rule which lowers the final accent in an IP. Under our analysis, these results provide direct support for our theory laid out in section 3.2.3. The asymmetry between the marking of (*restricted*) kontrast accents in theme-rheme (BA) and rheme-theme (AB) order is predicted by the asymmetry in the marking of pre- and post-nuclear prominence, without the need for a seemingly arbitrary rule of 'final lowering'. Liberman & Pierrehumbert also measured the scaling of the low following each accentual peak. They found that $f0$ fell more after Answer accents than Background accents. The authors claim the fall after the A accent is a phrase boundary, independent of the accent. Once more, we would say that it is correlated with the marking of nuclear prominence at the higher phrase level. We return to this in discussing our similar findings in Experiment 4 below.

From these studies, we can see that elements that are Contrastive, i.e. explicitly contrast with an equivalent element in the context, are realised with higher, and possibly later, $f0$ peaks and/or lower preceding $f0$ valleys. In many accounts this raising is identified as marking with L+H*. However, given the evidence from studies reported that looked at the acoustic correlates of these accents, and our discussion in section 3.2.4, it seems more likely that this marking is in fact increased prominence in general. That is, it evokes a *restricted*

kontrast interpretation. As discussed in section 2.2.3, this still leaves open the possibility that kontrastive themes are marked with a categorically distinct accent to kontrastive rhemes, e.g. L+H*(LH%) versus H*(LL%). Since most themes are unaccented, Contrastive contexts are often used to elicit accented themes. As we have just seen, however, the corresponding rhemes are not always Contrastive, and so the theme/rheme, contrastive/non-contrastive comparisons are conflated. In the last reported experiment, we saw that when Contrastive themes are compared with Contrastive rhemes, the theme accents seem to be lower than rheme accents (though the authors in that study were not actually investigating this comparison).

4.2 Experiments on the Nature of Theme and Rheme Accents

Much of the experimental work just reported conflates the prosodic marking of Contrastiveness, and theme/rheme status. However, it is still tenable that, when these are separated, kontrastive themes are distinguished from kontrastive rhemes by pitch accent or boundary tone type, as claimed in the literature. We saw in section 2.2.3 that Steedman (2000) claims the distinction is between L+H* and H*, while Büring (submitted) claims it is the whole tune (L+)H* LH% and H* LL% respectively. In section 3.2.3, we advanced our theory that theme/rheme status is signalled by relative prominence. Here we begin by trying to prove the opposite viewpoint. That is, in contexts where both the theme and the rheme are Contrastive (removing the confound), theme/rheme status is marked by the intonation contour, e.g. ToBI pitch accent and boundary tone type. This claim is tested in Experiments 1 and 2, complementary production and perception experiments. The results of these experiments did not support this hypothesis. Rather, they seemed to argue for our view that peaks in thematic accents are consistently lower than peaks in rhematic accents, something that is not easily captured in terms of ToBI intonation event type. In light of this finding, we pursue the nature of this relative height difference in Experiment 3, a reanalysis of our original experimental materials, and a second production experiment reported in Experiment 4, both of which further support our analysis.

4.2.1 Experiment 1: Production Experiment

The aim of the first experiment was to test whether themes are consistently produced with a different intonation contour *type* to rhemes in Contrastive contexts, e.g.:

(4.12) (That's Henry Lambert), (not Henry Lombard)
H* LL% H* L+H* LH%
(*theme* *rheme*) (*rheme* *theme*)

We only looked at the marking of theme/rheme status on *Lombard/Lambert*, as these were in nuclear position in each phrase, avoiding any confound with differences in the realisation of pre-nuclear and nuclear accents (see section 3.1.2). As suggested in section 3.1.3.1, we expect meaningful differences in accent shape to be realised minimally on nuclear accents and following phrase boundaries. In section 3.1.3.2, we saw that the phonetic distinction between L+H* and H* is disputed. Therefore, we tested firstly whether there is a consistent phonetic difference in the realisation of theme and rheme accents; and secondly, whether this difference can be framed in terms of the distinction between L+H* and H*. Our hypotheses for the first experiment, a smaller production experiment, therefore, were:

1. There is a consistent difference between the scaling and alignment of $f0$ turning points relative to the stressed syllable of thematic (T) and rhematic (R) nuclear accents.
2. This difference is consistent with marking by the ToBI pitch accents L+H* and H* respectively.
3. Theme phrases are marked by rising boundaries (H- or LH%), whereas rheme phrases are marked by flat or falling boundaries (L- or LL%).

4.2.1.1 Method

In order to control the information and segmental context as much as possible, and make detailed phonetic measurements, we used read speech. Eight sentences were constructed. The target word was in nuclear position in both theme and rheme phrases. Each target word was at least two syllables long with the primary stress not on the last syllable, in order to separate the pitch accent from any following boundary tone. The accented syllable in each target word began and ended with a sonorant consonant, providing a continuous $f0$ signal and making the speech signal easier to segment using a spectrogram.

Each sentence was presented in four versions, so that each target word would appear as both a theme and a rheme in both clauses of each sentence:

(4.13) Q: That guy's Henry Lombard, I think?
A: That's Henry *Lambert*, not Henry *Lombard*.

(4.14) Q: That guy's Henry Lombard, I think?
A: That isn't Henry *Lombard*, it's Henry *Lambert*.

(4.15) Q: That guy's Henry Lambert, I think?
A: That's Henry *Lombard*, not Henry *Lambert*.

(4.16) Q: That guy's Henry Lambert, I think?
A: That isn't Henry *Lambert*, it's Henry *Lombard*.

These sentences were divided randomly into four blocks, so that there was not more than one version of each sentence in any block. They were presented to the speaker along with 24 distractor sentences, making four blocks of 14 sentences each, or 56 sentences in total (see Appendix A for full list). This made a potential 32 tokens of each of the T and R accents.

One speaker, an undergraduate at the University of Edinburgh with a 'standard' (Edinburgh) Scottish English accent, was used for her ability to consistently produce well-modulated, natural-sounding speech when reading aloud. In a sound-proofed recording studio, the author asked the speaker each question in turn and our speaker replied. The dialogues were recorded digitally.

The target words (e.g. *Lombard* and *Lambert* above) were then analysed using *xwaves* (Entropic-Research-Labs 1998). The author judged, by listening to the recording and looking at the pitch track, whether the accented syllable in each target word was associated with a definite pitch movement in which the $f0$ turning points could be clearly determined. If it was, then, using the audio, pitch track, wave form and spectrogram associated with each word, the author labelled the following points in each accent:

1. C0: the beginning of the consonant of the stressed syllable
2. V0: the beginning of the vowel of the stressed syllable
3. C1: the beginning of the consonant following the stressed vowel
4. V1: the beginning of the vowel following the stressed vowel
5. L: the pitch low point, or point where the pitch track begins to rise sharply, before the pitch accent
6. H: the pitch peak, or the turning point of the pitch track at the height of the pitch accent
7. T0: the f0 level at the intensity peak in the last syllable in the word before the target one
8. T1: the f0 level at the intensity peak in the syllable following the accented one (but before any boundary tone rise, if present)

		C0	L	V0	H	C1	V1	T0-T1
*f*0 (Hz)	T accent	166.8	183.7	177.7	227.3	217.6	166.4	8.1
	R accent	210.0	212.8	232.4	267.2	260.4	186.9	54.2
Time (secs)	T accent	-0.059	-0.001	0.000	0.097	0.084	0.209	-
	R accent	-0.059	-0.053	0.000	0.101	0.083	0.199	-

Table 4.2: Results from Experiment 1: shows the $f0$ values and times at key points marked in target words for theme (T) and rheme (R) accents. Note times are normalised relative to V0, which is taken to be 0 secs. T0-T1 is the *difference* in $f0$ before and after the accent. N = 14.

4.2.1.2 Results and Discussion

Of the 32 theme tokens, 7 were judged by the author to have been produced with a pitch accent in which these points could be clearly determined. 29 of the 32 rheme tokens met this criteria. This result in itself indicates it is not just a *restricted* kontrast interpretation which leads to themes being pitch accented. The themes were also given, and it may be that our speaker, in this case, accomodated them thus and therefore only weakly accented, or deaccented them. But as, in this study, we were concerned with the realisation of accent shape, these other productions were put aside. Each of the seven T pitch accent tokens was matched with its equivalent R pitch accent token, and the remainder of the R tokens were excluded from analysis. Equivalent tokens were taken to be the same word in the same clausal position (first clause or second clause), but with a different function (theme or rheme). So, for example, the equivalent token of *Lambert* in (4.13) would be *Lambert* in (4.16); and the equivalent token of *Lambert* in (4.14) would be (4.15).

It was noted that in the rhematic clause, e.g. *it's Henry Lambert*, the preceding material was usually de-accented, or produced with a weak accent on *it's*; whereas in the thematic clause, e.g. *not Henry Lombard* there was usually a strong accent on *not*, although the make-up of the experimental materials did not allow this to be measured precisely. We return to this point below.

Table 4.2 shows the results from Experiment 1, where the labels are as described above. Times are normalised relative to V0, which is taken to be 0 seconds. As can be seen, there do seem to be small, but distinct, alignment differences between the two accents. For the T accent, L is aligned with V0; whereas the R accent begins to rise earlier, at C0. This result is significant using a two-tailed paired t-test ($t = 3.66$, $d.f. = 6$, $p < 0.011$). H, however, seems

	L-, LL%	H-, LH%
Theme	15	17
Rheme	30	2
Total	45	19

Table 4.3: Distribution of boundary types (rising versus falling) by information status (theme/rheme) following all target words (including unaccented words) in Experiment 1.

to be aligned a short way into the next consonant for both accents.

The results also suggest that the difference between the two accents could be indicated by pitch height. Both L and H were produced with lower $f0$ for T accents than for R accents. These results only tended towards significance ($t = 1.63$, $d.f. = 6$, $p < 0.154$ and $t = 2.01$, $d.f. = 6$, $p < 0.091$ respectively using two-tailed paired t-tests), however the sample size was small. R accents also seemed to be followed by a dip in $f0$, to well below the starting $f0$ level; whereas the $f0$ level after a T accent seemed to return approximately to its starting point. This result was significant (for $T0 - T1$, $t = 2.96$, $d.f. = 6$, $p < 0.025$ using a two-tailed paired t-test).

In order to test the last hypothesis, the author judged whether the boundary following each theme or rheme token was produced with a rising boundary (H- or LH%), or a flat/falling boundary (L- or LL%). This was done on the basis of a visual inspection of the $f0$ track and listening to the stimuli. All tokens, including those where the theme or rheme was not produced with a clear accent, were included. As shown in Table 4.3, theme tokens were significantly more likely to be followed by rising boundaries than rheme tokens ($\chi^2 = 16.8$, $d.f. = 1$, $p < 0.0001$). However, this relationship was asymmetric. While rhemes were highly likely to be followed by falling boundaries, themes were equally likely to be followed by either a rising or falling boundary. In separate chi squared tests, it was shown that there was no effect of Order (theme/rheme versus rheme/theme) or Type (e.g. *Lombard/Lambert*) on the likelihood of a rising boundary.

The results of our production study show that in utterances where our speaker produced pitch accents with clear turning points in the expected places (i.e. on the head of the theme or rheme phrase), she consistently produced T and R accents differently. It is not clear, however, whether it is the alignment differences, the relative pitch levels, the boundary tone or a combination of these forming the overall contour that conveys the distinction between theme and rheme. The alignment differences could be indicative of an accent type distinction, i.e.

between L+H* and H*, although the cues are much more subtle than those distinguishing other ToBI accents. As we discussed in the last chapter, relative height differences are not well captured in ToBI, making the scaling of H a problematic basis for the distinction. The last result would seem to be consistent with Büring's (submitted) contention that a high accent plus rising boundary tone marks thematic status. On the other hand, Steedman (2006*b*) would claim that is because theme marking is conflated with the meaning of rising boundaries, i.e. marking the phrase as the 'hearer's supposition', in this context.

4.2.2 Experiment 2: Perception Experiment

The second experiment, a perception study, tested which of the hypothesised differences between T and R accents, if any, are perceptible. Listeners were presented with a forced-choice exercise. Subjects heard two versions of the dialogues outlined above, with the pitch accent on the theme having been altered, and were asked to choose which dialogue they thought was more natural-sounding. There were two main hypotheses:

1. Subjects would prefer dialogues in which the pitch accent on the theme was produced with a T accent to dialogues where the theme was produced with an R accent.

2. Subjects would prefer dialogues in which the pitch accent on the theme was produced when each of four parameters (Alignment, Height, Fall and Boundary) was in the 't' setting, rather than the 'r' setting.

4.2.2.1 Method

The recordings from the first experiment were used to generate the stimulus materials. Four sentence types were used (1.7, 1.11, 1.12 and 1.14 in Appendix A and their variants). The questions were played back as they were recorded. The pitch tracks of the answers were manipulated and resynthesised using *Praat* tools (Boersma & Weenink 2003). Firstly, the pitch track of the entire answer was stylised automatically so that it was represented visually by straight lines drawn between pitch points at key turning points (approximately 15 per utterance). Then the position of these pitch points was altered manually so that there was a point at relevant locations (C0, V0, H, T1, B0 and B1, see below) in the pitch accent on the theme.

A Praat script was then used to generate 16 versions of each sentence. Each version had its key pitch points altered so all possible combinations of each of the following four parameters in each of their two hypothesised settings ('t'-like and 'r'-like) were produced, see Figure 4.3. The sentence was then resynthesised with the altered pitch track using the PSOLA technique. Pitch values were decided on the basis of the production study. Ratios

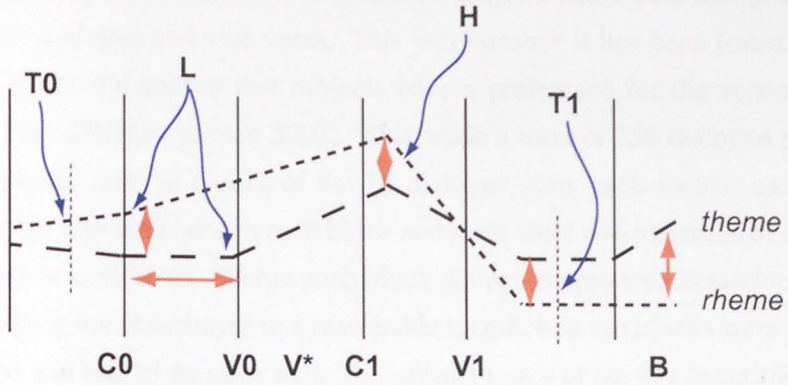


Figure 4.3: Manipulations of pitch accent shape used to create stimuli for Experiment 2. Manipulations were based on the productions of theme and rheme accents from Experiment 1. They included: *alignment*, of L relative to the onset of the stressed syllable (C0 versus V0); *height*, $f0$ value of L and H (varied simultaneously); *fall*, $f0$ of T1 relative to T0; and *boundary*, $f0$ in region B rising or flat.

were used rather than absolute differences in $f0$ as this is closer to human perception of pitch (see Ladd 1996, chap.7).

1. Alignment: 't': set time of L at V0, 'r': set time of L at C0
2. Height: Set time of H 20% into following C. Set $f0$ of L to be 20% less than H. 't': set H to be 210Hz, 'r': set H to be 250Hz
3. Fall: Set time of T1 at a stable point in the vowel following the accented one. 't': set $f0$ of T1 to be 10% lower than L, 'r': set $f0$ of T1 to be 20% lower than L
4. Boundary: Set time of B0 50ms before end of phrase, $f0$ same as T1. Set time of B1 at end of phrase. 't': set B1 to be 20% higher than B0, 'r': set B1 to be the same as B0

These answers were then used to set up pairs of dialogues for subjects to choose between. For the first hypothesis we paired answers that differed by three parameter settings (i.e. either 4 't'-like versus 1 't'-like (4-1) or 3 't'-like versus 0 't'-like (3-0), assuming that these were equivalent). The second hypothesis was tested by pairing answers that differed only by each one of the four parameters in turn (i.e. either 3-2 or 2-1).

Both versions ('It isn't X, it's Y' (theme-rheme) and 'It's Y, not X' (rheme-theme)) of each of the four answers were tested with each of the 16 resulting parameter pairings. In

addition, each pairing was tested in both orders, so subjects heard both the 'good' dialogue before the 'bad' dialogue and vice versa. This was because it has been found in previous forced-choice intonation studies that subjects have a preference for the version they hear most recently (e.g. Chorianopoulou 2002). This made a total of 256 dialogue pairs. These were divided evenly into 16 blocks of the 16 dialogue pairs each so that each of the 16 parameter settings appeared once in each block and there were two instances of each version of each dialogue in each block. Within each block dialogues appeared in random order.

In order to keep the experiment to a reasonable length, half the blocks were presented to half the subjects and half to the other half. Therefore, in each of the five experiments (one for hypothesis 1, four for hypothesis 2), subjects was the random factor. There were four within-subjects factors: Sentence (Lombard, London, Malaya, Wombats); Place (theme-rheme or rheme-theme); Order (good-bad or bad-good); and Type (the parameter pairing used - four combinations in hypothesis 1 and three in hypothesis 2).

Thirty subjects, staff and students at the University of Edinburgh, took part in the experiment in return for a small monetary reward. Subjects were told that they would hear two dialogues, and that the intonation contour of the answer would be different in each one. They were asked to choose which answer sounded like a more natural response to the question. Subjects began with a practice block consisting of 16 4-1 and 3-0 sentences not in the main experiment. They then heard eight blocks of dialogues, with a break after every two blocks. Subjects were told there was no time pressure in responding but that the dialogues could not be repeated. The entire session took about 45 minutes.

4.2.2.2 Results

In relation to the first hypothesis, it was found that subjects did prefer answers produced with a T accent on the theme to answers with an R accent on the theme. Overall 66.7% chose the 4-1 and 3-0 sentences with more 't' settings. This was significantly more than chance (χ^2 = 115.8, $d.f.$ = 1, $p < 0.01$). However, this result was affected both by the Order in which the stimuli were presented ('good'-'bad' and 'bad'-'good') and the Place of the theme accent (theme-rheme and rheme-theme). Using a 1 x 2 repeated-measures ANOVA there was a significant main effect of Order, $F(1,24) = 6.508$, $p < 0.018$; similarly for Place, $F(1,24) = 4.617$, $p < 0.042$. These two variables seemed to interact, though not significantly, $F(1,24) = 2.062$, $p < 0.164$. This can be seen in Figure 4.4. For the theme-rheme ordered answers, the Order was significant. When subjects heard the 'good' (more theme-like) version second, they preferred it 66.9% of the time, whereas when the good version was presented first, they performed only at the level of chance. For the rheme-theme ordered sentences, however, subjects reliably preferred the more theme-like version in either order.

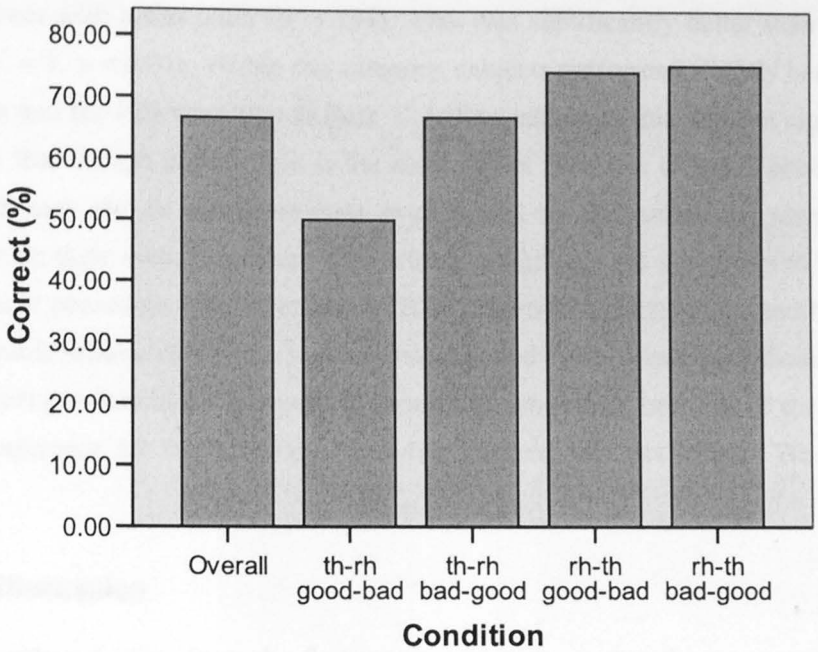


Figure 4.4: Results for Hypothesis 1 in Experiment 3. Shows the proportion of times subjects preferred a *theme*-like accent to a *rheme*-like accent on a theme target overall, as well as broken down by Place and Order. This demonstrates that in utterances where the theme target word followed the rheme (i.e. *it's Y, not X*), subjects preferred *theme*-like accents in either presentation order. However, when the theme target preceded the rheme (i.e. *it isn't X, it's Y*), subjects only performed above chance when they heard the "good" version of the utterance (i.e. with a *theme*-like accent on the theme target) before the "bad" version. N = 224.

Subjects performed better on the *Lombard-Lambert* sentences (1.14 and variants in Appendix A), and worse on the *monkeys-wombats* sentences (1.11 and variants in Appendix A), 82.3% and 56.5% respectively, although there was no significant main effect of Sentence ($F(3,22) = 1.195$, $p = 0.335$). Many subjects commented on the strangeness of the *monkeys-wombats* sentences, suggesting that the preference for T accents on themes in appropriate and familiar contexts is even stronger than these results suggest. Our assumption that inverse parameter settings could be treated as equal (e.g. that 4-1 is effectively the same as 3-0) proved to be justified, 67.4% and 65.9% respectively; there is no main effect of Type.

In relation to the second hypothesis, the only single 't' parameter setting which caused subjects to significantly prefer that answer was Height. 73.4% of subjects chose the versions

of the answer with lower pitch (N = 334). This was significantly better than chance (χ^2 = 140.8, $d.f.$ = 1, $p < 0.01$). Within this category, subjects performed slightly better if both the Alignment and the Fall were also in their 't' setting, although this was not significant. This may show that though pitch height is the most robust indicator of the T accent, alignment and the fall may also be secondary cues, even if they are not sufficiently strong to indicate the accent on their own. Boundary type (rising or falling) did not prove to be significant at all. This is consistent with Steedman's (2000) claim that pitch accent and boundary tone have separable influences on information structure, and that accent type indicates themehood (contra Büring submitted). However, it should be noted that, because of the experimental design, preference for the boundary following a *rheme* was not tested. We return to this below.

4.2.2.3 Discussion

The most robust finding from the first two experiments is that theme accents are realised with lower peaks than rheme accents. Could this be accommodated in the ToBI distinction between L+H* and H*, as was hypothesised above? As we saw in section 4.1, previous work on the realisation of Contrastive accents compared to non-Contrastive accents, analysed in terms of L+H* and H*, showed the main difference was that peaks in L+H* accents are *higher*. However, our work here indicates that peaks in thematic accents, also analysed as L+H*, are *lower*. This would seem to point to two diametrically opposed uses for the L+H* accent. Further, the scaling and alignment distinctions on the beginning of the accent rise (L) found in the production experiment were not significant in the perception experiment. In any case, as discussed in section 3.1.3.2, these differences are much more subtle than in the rest of the ToBI system, which describes association between tones and syllables, not parts of syllables. All of these points would seem to argue for the contention in Ladd & Schepman (2003) and Dilley (2005) that L+H* and H* should be merged in the description of English intonation.

Another possibility would be to argue that themes are *downstepped* relative to rhemes. That is, themes are !H* in nuclear position, while rhemes are H*. However, this could lead to a logical difficulty that a Contrastive theme in the same phrase as a non-Contrastive rheme would be both downstepped and raised for emphasis. It is also difficult to resolve with evidence that !H* is interpreted as rhematic (cf. Ayers 1996), but marking the referent as relatively more accessible (cf. Baumann 2005) (see sections 3.1.2 and 2.2.2.2 respectively). A further problem comes with the ordering of elements. Phrases can be in theme-rheme, rheme-theme, or even rheme-rheme order (the last in an all-new sentence). However, accents cannot be pre-downstepped, e.g. a !H* H* sequence, so the status of the first accent in any

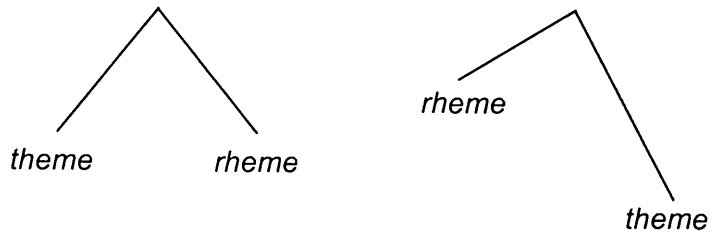


Figure 4.5: Diagrammatic representation of the signalling of the relative metrical prominence of theme and rheme nuclear accents.

phrase would be indeterminate. Phrases containing only one accent would also be inherently ambiguous.

Conceptually, however, the biggest difficulty with this is that it misses the *relationship* inherent in the peak height distinction. Themes are semantically subordinate to rhemes. They 'set up' the context in which the rheme is resolved. Under our theory, set out in section 3.2.3, this notion is directly captured by the relative prominence relationship. We claimed that, when the theme and the rheme form different phrases, their status is indicated by relative prominence at the level of phrasing that includes both, i.e. rhemes are nuclear. As we set out there, the phonetic signalling of this relationship mirrors that at the phrase level, i.e. the last of roughly equally acoustically prominent accents is perceived as nuclear over the larger phrase, while post-nuclear accents must be considerably less acoustically prominent. This yields the theme/rheme relationship shown in Figure 4.5 (repeated from Figure 3.12). Our theory could explain the presentation order effect found in the perception study. Subjects were sensitive to the peak height distinction in rheme-theme order, but were not nearly so reliable in theme-rheme order (Figure 4.4). Given our explanation, this would be expected. Because of the right-branching bias, in rheme-theme order the theme accent must be clearly lower. In theme-rheme order, the ordering of the accents is enough for the right-most accent to be perceived as nuclear.[2] This could also be affected by the greater frequency of theme-rheme order in language overall (see results in section 6.4), making it is less important that it be marked prosodically. Hence subjects' difficulty in this condition.

In the first experiment, we found that rheme accents end with a 'dip' in $f0$ (T0-T1), and are almost always followed by a falling boundary. As noted above, given the design of the

[2]Note, however, that this is theme/rheme status over the whole utterance, not within each clause. So this effect could also mark the Contrastive relationship between these elements across phrases. We return to this below.

current experiment, we did not test the acceptability of these variables after *rheme* accents. It may be that this fall and/or the low boundary, is an important cue to the perception of a rheme, but does not actively interfere with the acceptability of a theme. In section 3.1.2, it was noted that nuclear accents are often followed by an $f0$ fall, so this would be consistent with the marking of the nuclear status of the rheme over the higher prosodic phrase. Similarly, the falling boundary could mark the close of the information unit.

It will be recalled that, in Experiment 1, the height difference between theme and rheme accents only tended toward significance. However, most of the theme accents were excluded because their $f0$ turning points could not be reliably determined. Having rejected the accent shape hypothesis, we need to look at this experimental material again. Our theory is that themes are prosodically subordinate to their paired rhemes. In *That's Henry Lambert, not Henry Lombard*, the paired theme of *Lambert* is *that's*, and the paired rheme of *Lombard* is *not* (see (4.12)). As noted above, contexts such as *that's* tended to be deaccented or weakly accented, and contexts such as *not* tended to be strongly accented. This is therefore further support for our theory: in most cases, themes like *Lombard* were either deaccented or weakly accented to mark them as prosodically subordinate to their paired rheme, e.g. *not*. Unfortunately, as contextual material was not controlled in the stimuli make-up, it is was not possible to accurately measure the relative heights of theme and rheme pairs to test our new hypothesis. In Experiment 3, however, we reanalyse the material from Experiment 1 to show that the excluded theme accents are significantly lower than rheme accents in the original material. In Experiment 4 we use a new set of materials that allow us to measure the relative heights of theme/rheme pairs directly. Further, we consider another possible interpretation of the relative height distinction found in Experiments 1 and 2. In all of experimental stimuli, e.g. *Lombard/Lambert*, the target theme and rheme tokens were in a Contrastive relationship. In section 3.2.3, we suggested that Contrastive relationships themselves could be conveyed by relative accent height across phrases. This possibility is tested in Experiments 3 and 4.

Lastly, there is an intuitive sense in which our explanation does not capture the whole story. Impressionistically, stimuli in the second experiment with a lower accent on the theme but all other parameters in the 'r'-setting did not 'sound' right. Following from the discussion in section 3.3.2, it may be that the subtle alignment and scaling differences on the L in the first experiment were meaningful at another level of linguistic structure. We know that these are not interpreted categorically (see Ladd & Schepman 2003, Dilley 2005), yet they may still convey shades of meaning. A number of subjects volunteered that, for some stimuli in the second experiment, the speaker sounded more 'annoyed' or 'irritated' in one case than the other. It could be that the later L had a 'softening' effect on the contradiction implied by the stimuli. We did not directly test this sort of explanation, which indeed would be extremely

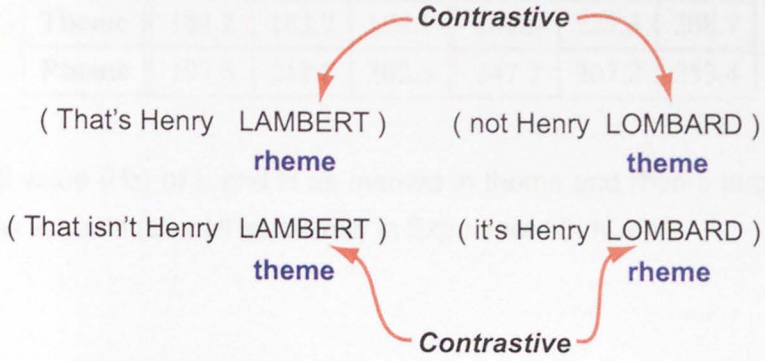


Figure 4.6: The relationship between Contrastive theme and rheme accents used to test Hypothesis 2 in Experiment 3.

difficult. However, we return to this in Chapter 7, arguing that part of the reason thematic kontrast has been thought to be signalled by pitch accent type is because of the illocutionary and affective connotations correlated with it.

4.2.3 Experiment 3: Relative Height

Our new hypothesis is that theme accents are less relatively prominent than rheme accents within metrical prosodic structure. That is, the prominence distinction is shown by ordering in theme-rheme order, and peak height in rheme-theme order. As just discussed, the peak height distinction in the first experiment only approached significance. However, in that experiment we were interested in accent shape, i.e. $f0$ alignment and scaling. Therefore a large number of the theme productions were not included in the analysis because their $f0$ turning points could not be accurately determined. If, on the other hand, the relevant distinction is indicated by relative prominence, or accent strength, and not accent shape, then these productions are obviously directly relevant. In this experiment, therefore, we went back and examined the scaling of L and H on each theme and rheme target including the 'weak' accents. Completely deaccented cases were still excluded, however.

At the end of the last section, we discussed why the material relative height distinction is not in fact between theme and rheme accents per se, but between paired theme and rheme accents, i.e. in the same information unit. As we discussed there, the materials from Experiment 1 do not allow us to test this directly, however we will look at this in Experiment 4. It

	L			H		
	Weak	**Full**	**All**	**Weak**	**Full**	**All**
Theme	184.2	183.7	184.1	200.5	227.3	208.7
Rheme	197.8	212.8	202.3	247.3	267.2	253.4

Table 4.4: $f0$ value (Hz) of L and H as marked in theme and rheme target words with Full and Weak accents (as well as overall) in Experiment 3. N = 46.

was also suggested that there could be a prosodically weak-strong relationship between Contrastive theme and rheme tokens, as illustrated in Figure 4.6. In this experiment, therefore, we tested whether there is a correlation between the heights of these accents.

Therefore, the hypotheses in the third experiment were:

1. Accents on themes are consistently scaled lower than accents on rhemes.
2. There is a correlation between the heights of Contrastive theme and rheme accents (i.e. in the same stimulus).

4.2.3.1 Method

The target words from Experiment 1 were used, with further measurements to test the current hypotheses. Praat was used for all analyses. In order to test the first hypothesis, the prominences on themes were classified as *Full*, i.e. a well-defined accentual pitch movement included in the first experiment, *Weak*, i.e. any $f0$ movement on the syllable with primary lexical stress in the target word; or *Deaccented*, i.e. a flat $f0$ track (see Figure 4.7).

For all Weak accents, the Low (L) preceding the accentual rise, and the High (H) at the $f0$ peak were marked as in the first experiments and the values for all theme tokens recorded. As can be seen in Figure 4.7, with the weak accents the $f0$ level before the accentual rise and the $f0$ height of the accent can be found with reasonable accuracy. However, the precise turning points cannot be determined as clearly, hence their exclusion from the first experiment. L and H points in the equivalent rheme tokens were also taken (as defined in the first experiment). To test the second hypothesis, we compared the scaling of L and H for Contrastive theme and rheme tokens in the same stimulus (re Figure 4.6).

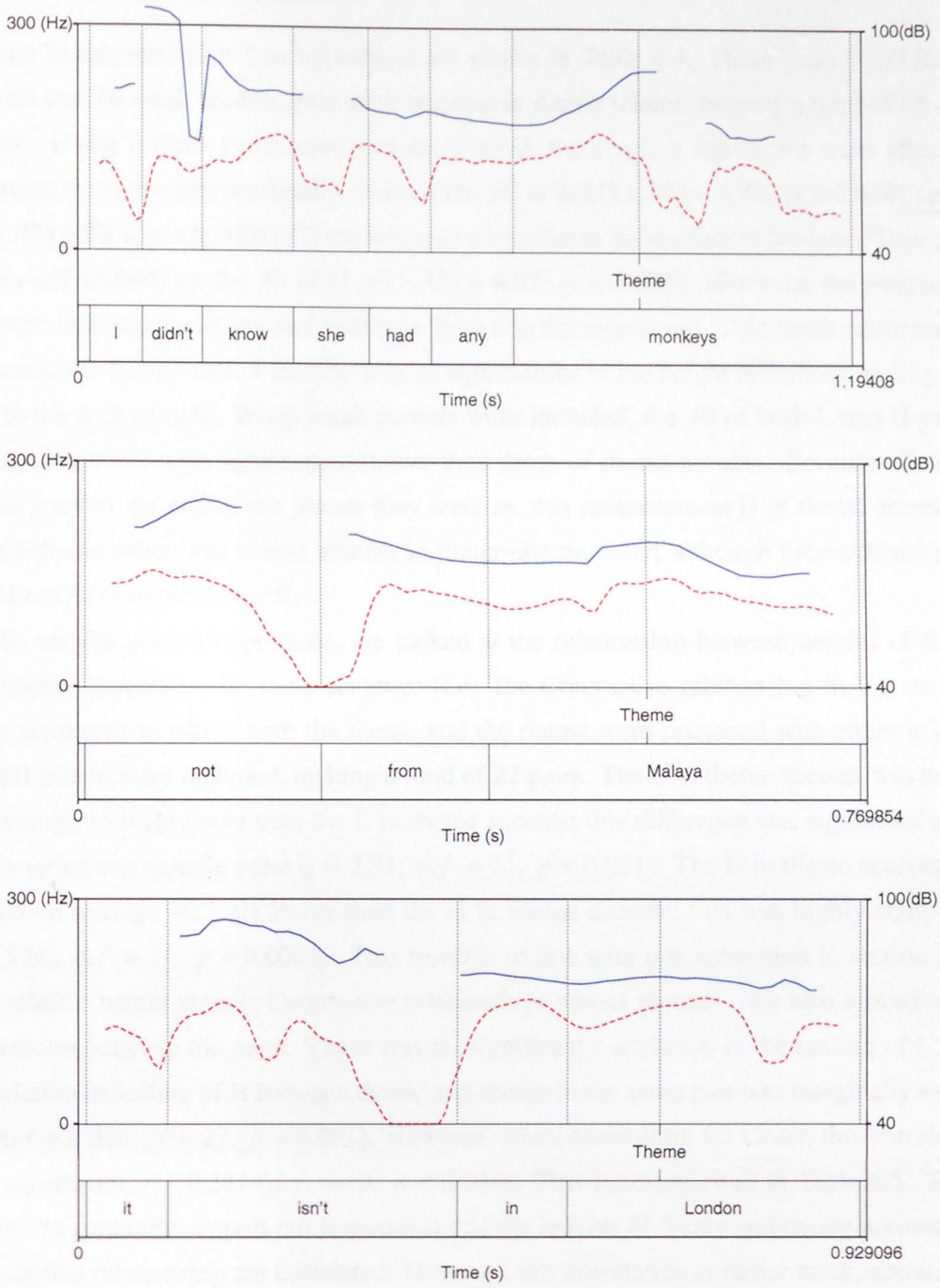


Figure 4.7: Examples of prominences on themes from the first production experiment categorised as Full accent, Weak accent and Deaccented (blue line is the $f0$ track, dashed red line the intensity curve).

4.2.3.2 Results and Discussion

Results in relation to the first hypothesis are shown in Table 4.4. There were 7 full theme accents and 16 weak accents, plus their equivalent rheme tokens, making a total of 46 data points. Using a three factor multivariate ANOVA there was a significant main effect of Information Status (theme/rheme) on both the $f0$ of L ($F(1,42) = 6.58$, $p < 0.018$) and H ($F(1,42) = 29.4$, $p < 0.0001$). There was also a significant main effect of Sentence Type (e.g. *Lombard/Lambert*) on the $f0$ of H ($F(1,42) = 4.67$, $p < 0.002$). However, the interaction between Information Status and Sentence Type was not significant. This result confirms our contention in Experiment 1 that the lack of significance in the height difference finding was due to the lack of data. When weak accents were included, the $f0$ of both L and H points of theme accents were significantly lower than those of rheme accents. Because all target words were at the end of the phrase they were in, this measurement is of theme accents in rheme-theme order, and rheme accents in theme-rheme order; although theme/rheme pairs could not be compared directly.

To test the second hypothesis, we looked at the relationship between heights of theme and rheme accents in the same sentence (i.e. the Contrastive relationship in Figure 4.6). Only sentences in which both the theme and the rheme were produced with either a Weak or Full accent were included, making a total of 22 pairs. The L in theme accents was scaled on average 15.0 Hz lower than the L in rheme accents; this difference was significant using a two-tailed one-sample t-test ($t = 2.31$, $d.f. = 21$, $p < 0.031$). The H in theme accents was scaled on average 40.1 Hz lower than the H in rheme accents; this was highly significant ($t = 5.66$, $d.f. = 21$, $p < 0.0001$). This result is in line with our contention in section 3.2.3 that relative height signals Contrastive relationships across phrases. We also looked at the correlation between the pairs. There was no significant correlation in the scaling of L. The correlation in scaling of H between theme and rheme in the same pair was marginally significant, $r = 0.358$ ($N = 22$, $p < 0.051$). However, when controlling for Order, this correlation was significant, $r = 0.381$ ($d.f. = 19$, $p < 0.044$). This is summarised in Table 4.5. These results do generally support our hypothesis that the heights of theme and rheme accents in a Contrastive relationship are correlated. However, this correlation is rather weak, accounting for only 15% of the variance in the scaling of H between themes and rhemes.

The results from Experiment 3 do generally confirm our analysis of the theme-rheme relationship, advanced in the last chapter. That is, themes are prosodically subordinate to rhemes at the level of phrasing that includes both. When all tokens were taken into account, theme accents were scaled lower than rheme accents in rheme-theme order. However, given the set-up of the materials, it was not possible to compare the heights of paired theme and rheme tokens. Experiment 4 looks at this directly. We also examine whether the theme-

	Mean difference (Hz)	s.d.	Corr.	Var. (R^2)
Contrasts	40.1	33.2	0.381	14.5%

Table 4.5: Mean difference and standard deviation (s.d.) in the scaling of H in theme and rheme accents in a Contrastive relationship in Experiment 3; as well as the correlation (controlling for order) between these accents, and the percentage of variance (R^2) in the scaling of H for these accents accounted for by this correlation. N = 22.

rheme and the Contrastive relationships are both marked prosodically, and which is stronger.

4.2.4 Experiment 4: Production Experiment

The design of Experiment 4 was similar to Experiment 1, except that the materials were changed to allow a direct comparison of the heights of theme and rheme accents within the same phrase. As just set out, the results from the first three experiments were consistent with our theory that themes are prosodically subordinate to rhemes. However, because of the stimulus design, we could not test whether themes are lower than rhemes in some absolute sense, or whether it is the prosodic subordination *relationship* between theme and rheme pairs that is important. Finally, the earlier materials did not allow us to compare this relationship to any relative height distinction between Contrastive elements, which earlier results also suggest exists. We therefore carried out another small production study with a set of six sentences such as (4.17) where theme and rheme status was systematically varied both within and across clauses (see Appendix B for the full list).

(4.17) Q: You're going to see Amanda tomorrow, right?
A: No, [I'm seeing Amanda] [on Monday],
theme *rheme*
[I'll see] [Norma] [tomorrow].
rheme *theme*

As is indicated, within each clause there is a theme-rheme relationship between *Amanda* and *Monday*, and *Norma* and *tomorrow* respectively; given the question (note that square brackets indicate information structure boundaries, not prosodic boundaries). There is also a Contrastive relationship between *Monday* and *tomorrow*, and between *Amanda* and *Norma*. We would expect the relationship between the Contrastive theme and rheme to also be weak-strong like the *Lombard/Lambert* relationship above.

We tested the following hypotheses:

1. The $f0$ turning points (L and H) of thematic accents are consistently scaled lower than those of rhematic accents.

2. The dip (T0-T1) following rhematic accents is greater than following thematic accents.

3. The peaks (H) of thematic accents are scaled consistently lower than the peaks of their rhematic accent pair. This effect is stronger in rheme/theme order than theme/rheme order.

4. The peaks (H) of thematic accents are scaled consistently lower than the peaks of rhematic accents they are in a Contrastive relationship with. This effect is stronger in rheme/theme order than theme/rheme order.

4.2.4.1 Method and Results

The experimental method was similar to the first experiment, except this time a male American English speaker was used. He produced a total of 67 theme accents and 85 rheme accents according to the judgement of the author. Measurements of L and H relative to C0, V0 and C1 were taken, as well as T0-T1, as defined in the first experiment. Results, showing the values for the first and second clause (e.g. *I'm seeing Amanda on Monday* and *I'll see Norma tomorrow* respectively in (4.17)) are shown in Table 4.6, as well the two clauses combined in Table 4.7. Since $f0$ always declined, results for the accents in first position (e.g. *Amanda* and *Norma*) are given separately to those in second position (e.g. *Monday* and *tomorrow*).

Results supported the first hypothesis, although the effect of phrasal position can be seen. Using a three factor multivariate ANOVA, there was a significant main effect of Information Status (theme/rheme) on the $f0$ of L ($F = 8.37$, $d.f. = 1$, $p < 0.004$) and the $f0$ of H ($F = 18.22$, $d.f. = 1$, $p < 0.0001$). There was also a significant main effect of Position (first/second) on the $f0$ of L and H. There was no significant effect of Clause (first/second). There was further a significant effect of Info*Position on the $f0$ of L ($F = 18.49$, $d.f. = 1$, $p < 0.0001$) and H ($F = 4.71$, $d.f. = 1$, $p < 0.032$). This can be seen in Table 4.7: there is very little difference in the height of L and H at the beginning of the phrase, whereas at the end of the phrase L and H in theme accents were significantly lower than in rheme accents. There was no significant effect of Information Status on the alignment of either L or H, supporting our contention that the alignment difference found in Experiment 1 was not robust.

Results also supported the second hypothesis. There was a significant main effect of Information Status on T0-T1 ($F = 14.2$, $d.f. = 1$, $p < 0.0001$). There was also a significant

First Clause								
		Accents in First Position						
		C0	L	V0	H	C1	V1	T0-T1
*f*0 (Hz)	T accent	152.8	151.5	166.3	190.7	189.4	176.3	9.41
	R accent	148.8	147.6	166.4	197.7	196.4	175.5	28.9
Time (secs)	T accent	-0.073	-0.041	0	0.091	0.094	0.164	-
	R accent	-0.086	-0.043	0	0.098	0.101	0.182	-
		Accents in Second Position						
		C0	L	V0	H	C1	V1	T0-T1
*f*0 (Hz)	T accent	123.7	118.8	134.1	144.3	142.1	131.1	15.6
	R accent	129.5	131.6	148.4	167.7	166.5	147.6	48.8
Time (secs)	T accent	-0.174	-0.058	0	0.090	0.106	0.185	-
	R accent	-0.077	-0.051	0	0.087	0.101	0.174	-
Second Clause								
		Accents in First Position						
		C0	L	V0	H	C1	V1	T0-T1
*f*0 (Hz)	T accent	126.5	128.0	146.0	179.1	176.9	150.3	15.7
	R accent	125.8	127.1	150.1	184.5	179.8	152.3	26.6
Time (secs)	T accent	-0.078	-0.043	0	0.098	0.102	0.183	-
	R accent	-0.083	-0.048	0	0.095	0.100	0.182	-
		Accents in Second Position						
		C0	L	V0	H	C1	V1	T0-T1
*f*0 (Hz)	T accent	105.8	107.7	115.5	128.5	127.7	117.6	30.2
	R accent	113.9	119.0	127.2	143.3	142.0	129.4	43.8
Time (secs)	T accent	-0.099	-0.037	0	0.088	0.099	0.172	-
	R accent	-0.069	-0.030	0	0.095	0.112	0.179	-

Table 4.6: Results from Experiment 4: shows the $f0$ values and times at key points marked in target words for theme (T) and rheme (R) accents. Note times are normalised relative to V0, which is taken to be 0 secs. T0-T1 is the *difference* in $f0$ before and after the accent. Results for each clause (e.g. *I'm seeing Amanda...* versus *I'll see Norma...* in (4.17)), and each position in each clause (e.g. *Amanda/Norma* versus *Monday/tomorrow*) are shown separately. N = 152.

Clauses Combined								
		Accents in First Position						
		C0	**L**	**V0**	**H**	**C1**	**V1**	**T0-T1**
***f*0 (Hz)**	**T accent**	140.6	140.6	156.9	185.3	183.6	164.2	12.4
	R accent	136.5	136.7	157.7	190.7	187.6	163.1	27.7
Time (secs)	**T accent**	-0.075	-0.042	0	0.095	0.098	0.173	-
	R accent	-0.084	-0.046	0	0.096	0.101	0.182	-
		Accents in Second Position						
		C0	**L**	**V0**	**H**	**C1**	**V1**	**T0-T1**
***f*0 (Hz)**	**T accent**	111.0	110.9	120.9	133.1	131.8	121.5	25.9
	R accent	122.6	126.0	138.9	156.8	155.6	139.5	46.5
Time (secs)	**T accent**	-0.097	-0.043	0	0.089	0.101	0.176	-
	R accent	-0.074	-0.041	0	0.091	0.106	0.176	-

Table 4.7: Results from Experiment 4: as in Table 4.6, shows the $f0$ values and times at key points marked in target words for theme (T) and rheme (R) accents. Note times are normalised relative to V0, which is taken to be 0 secs. T0-T1 is the *difference* in $f0$ before and after the accent. Results for both clauses (e.g. *I'm seeing Amanda...* and *I'll see Norma...* in (4.17)) are collapsed, so only each position (e.g. *Amanda/Norma* versus *Monday/tomorrow*) is shown separately. N = 152.

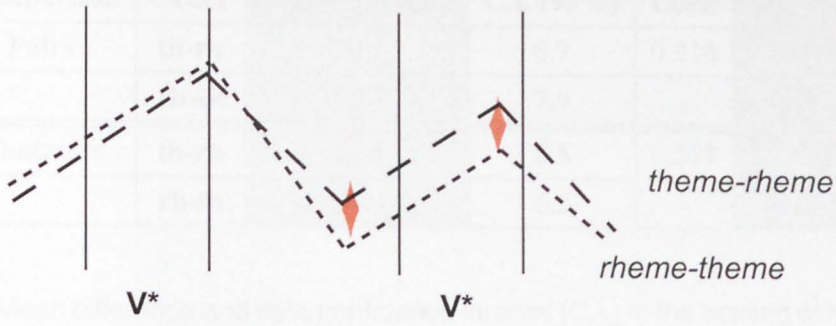


Figure 4.8: Diagrammatic representation of the results from Experiment 4, showing the significant difference in the depth of the fall in $f0$ following the first accent (e.g. after *Norma/Amanda*), and in the height in $f0$ of the second accent (e.g. on *Monday/tomorrow*) by order (i.e. theme-rheme versus rheme-theme).

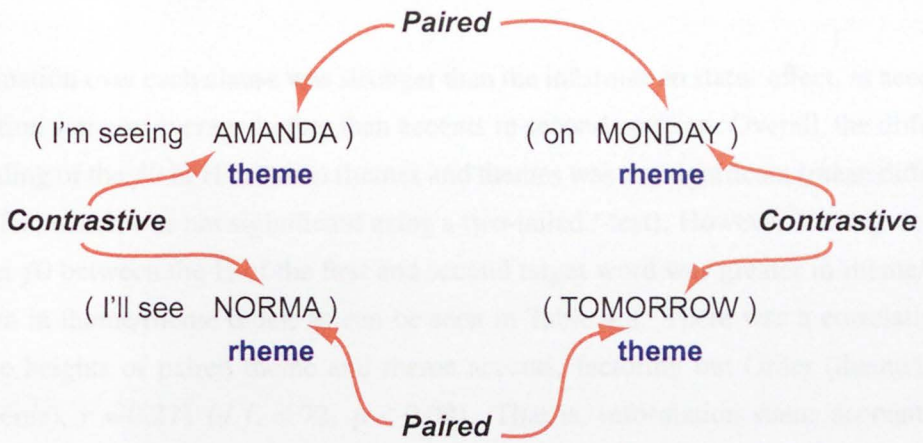


Figure 4.9: Illustration of Paired (Hypothesis 3) and Contrastive (Hypothesis 4) comparisons between theme and rheme accents made in Experiment 4.

main effect of Position on T0-T1 ($F = 7.96$, $d.f. = 1$, $p < 0.005$). However, the interaction Info*Position was not significant. Over both positions there is a greater drop in $f0$ following rheme accents than theme accents. In second position the drop is greater than in first position in both clauses (see Table 4.6). The resulting overall contours can be seen in Figure 4.8.

Our third hypothesis was that themes would always be lower than their paired rhemes, but that this effect would be stronger in rheme-theme order than in theme-rheme order. Theme-rheme pairs are target words in the same clause, as shown in Figure 4.9. In fact, it turned out

Comparison	Order	Mean diff (Hz)	C.I. (95%)	Corr.	Var. (R^2)
Pairs	**th-rh**	29.8	6.7	0.316	10.0%
	rh-th	59.3	7.9	-	not sig.
Contrasts	**th-rh**	-4.4	6.5	0.579	33.5%
	rh-th	28.4	6.5	-	not sig.

Table 4.8: Mean difference and 95% confidence interval (C.I.) in the scaling of H in Paired and Contrastive theme and rheme accents by order (theme-rheme versus rheme-theme) in Experiment 4; as well as the correlation (controlling for order) between these accents, and the percentage of variance (R^2) in the scaling of H for these accents accounted for by these correlations respectively. Note correlations were not significant ($p > 0.05$) in rheme-theme order. N = 152.

that declination over each clause was stronger than the information status effect, as accents in first position were on average higher than accents in second position. Overall, the difference in the scaling of the $f0$ of H between rhemes and themes was not significant (mean difference was 4.61Hz, which was not siginificant using a two-tailed t-test). However, the degree of the decline in $f0$ between the H of the first and second target word was greater in rheme/theme order than in theme/rheme order, as can be seen in Table 4.8. There was a correlation between the heights of paired theme and rheme accents, factoring out Order (theme/rheme, rheme/theme), $r = 0.271$ ($d.f. = 72$, $p < 0.02$). That is, information status accounted for 7.3% of the variation in $f0$ peak height. It is useful to break this down into the correlation between peak heights in each order. In theme/rheme order, there is a small but significant correlation between peak heights ($N = 46$, $p < 0.03$), see Table 4.8. In rheme/theme order, there is no significant correlation between peak heights, suggesting a greater disassociation in this case.

Our final hypothesis was that themes would be consistently lower than rhemes with which they were in a Contrastive relationship, see Figure 4.9. In this case, the effect of declination was not as strong, as both comparisons involved accents in the same position in each phrase. Overall, the $f0$ of H in themes was lower than in rhemes with which they were in a Contrastive relationship. The mean difference was 14.59 Hz, which was significant using a two-tailed t-test ($t = 4.96$, $d.f. = 75$, $p < 0.0001$). There was no significant effect on the scaling of L. Using a three factor multivariate ANOVA, there was a significant main effect of Order (theme/rheme, rheme/theme) on the difference in the $f0$ of H

($F = 52.98$, $d.f. = 1$, $p < 0.0001$). Once more, the relative height difference only shows up in rheme-theme order, as can be seen in Table 4.8. There was also a significant main effect of Position (first, second) on the difference in H ($F = 17.3$, $d.f. = 1$, $p < 0.0001$). Again, in theme/rheme order there was no significant difference in the height of the accents. In rheme/theme order, in first position (e.g. the *Amanda/Norma* comparison) the mean difference was 17.6 Hz, whereas in second position (e.g. the *Monday/tomorrow* comparison) the mean difference was 39.2 Hz. There was no significant main effect of Sentence Type (e.g. *Amanda*), nor of the interaction between Order and Position. Again we can see the asymmetry in the marking of a subordinate relationship prosodically: in weak-strong order, it is the right-branching prosodic structure which marks the second element as more prominent, even if it is not actually higher. In strong-weak order, the second element must be acoustically less prominent to be perceived as less prominent. Finally, we can again see a correlation between the heights of theme and rheme accents in a Contrastive relationship. Overall $r = 0.404$ ($d.f. = 72$, $p < 0.0001$) when factoring out Order and Position. Therefore, the marking of this relationship accounted for 16.3% of the variation in the heights of the respective accents. We again looked at the correlation in each order, factoring out Position. As can be seen in Table 4.8, there was a significant correlation in theme/rheme order ($d.f. = 29$, $p < 0.001$), but in rheme/rheme order, this was only marginally significant $r = 0.283$ ($d.f. = 41$, $p < 0.066$). These results, along with those for the paired relationship above, suggest that there is a much stronger association between peak heights in theme/rheme order than rheme/theme order, something our final model should try to explain.

4.2.4.2 Discussion

The results of Experiments 3 and 4 support our contention that the theme-rheme relationship is conveyed in terms of a weak-strong prosodic relationship, re Figure 4.5. In Experiment 4, we also showed that in theme/rheme order accent height is more strongly correlated than in rheme/theme order, suggesting a greater prosodic dissociation. These later two experiments further seem to show that the Contrastive relationship is conveyed by relative height across phrases. The correlation between accent heights is if anything stronger between theme and rheme accents in a Contrastive relationship than in a paired relationship. Can these findings be reconciled in terms of relative prominence structure, as we saw in the last chapter?

Figure 4.10 gives a possible branching structure consistent with the results from Experiment 4. As can be seen, in theme-rheme order, paired theme and rheme accents form part of the same phrase. However, the effect of having an accented theme following a rheme is to dissociate the two accents with a phrase break. If this break were not there, the theme would be deaccented (i.e. an accent with no pitch movement) in post-nuclear position. This allows

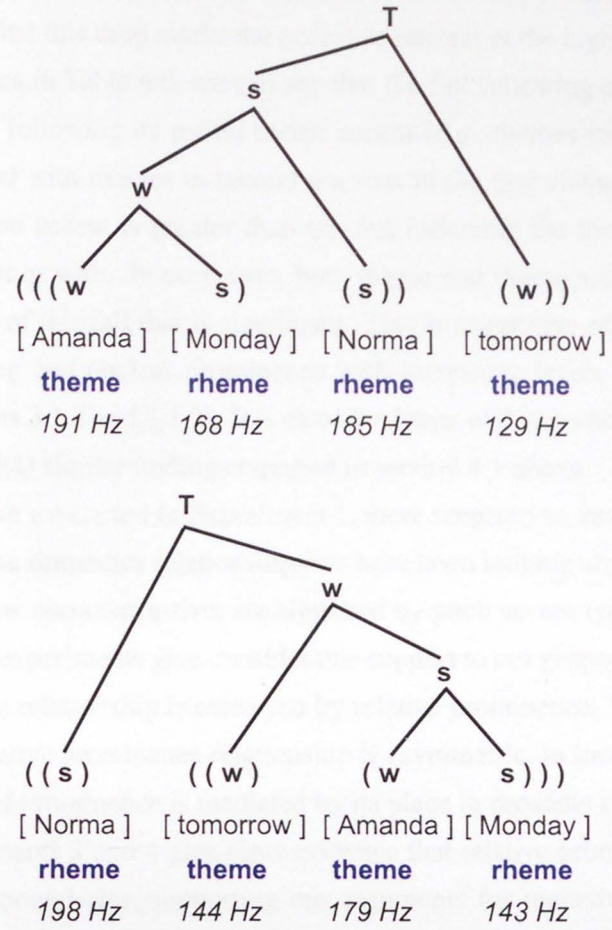


Figure 4.10: Possible metrical structure showing prominence relationships between accents on theme/rheme target words in Experiment 4. The rest of each utterance is omitted for clarity. Phrase structure is shown by parentheses. Theme/rheme status, along with the average $f0$ at the accent peak (H) for each target word are shown.

either the theme or rheme accent to be grouped more closely prosodically with surrounding material, including its Contrast, than with its paired rheme or theme accent, explaining the correlation results. For the Contrastive relationships, we again see each pair is grouped more closely in theme-rheme order than rheme-theme order, e.g. *Amanda* and *Norma* have one phrase boundary between them in the first ordering, but two in the second ordering. Again, this is supported by the correlation results. These tree structures also seem to reflect the mean $f0$ of H for each accent reported in Figure 4.10 (see further Table 4.6).

Like in Experiment 1, in Experiment 4 we found that the average fall in $f0$ level after

an accent (T0-T1) was greater with rheme accents than with theme accents. This supports our earlier analysis that this drop marks the accent as nuclear at the higher level of phrasing. Looking at the figures in Table 4.6, we can see that the fall following each rheme accent is greater than the fall following its paired theme accent (e.g. themes in first position in the first clause are paired with rhemes in second position in the first clause).[3] Further, the fall following each rheme accent is greater than the fall following the theme accent it is in a Contrastive relationship with. In each case, both theme and rheme accents have falls after them, it is the *depth* of this fall that is significant. This is suggestive of increasing phonetic cues to both phrasing and nuclear prominence with increasing levels of recursive phrasal structure (see sections 3.1.1 and 3.1.2). It is also consistent with our analysis of Liberman & Pierrehumbert's (1984) similar findings reported in section 4.1 above.

To return to where we started in Experiment 1, there seems to be very little evidence that either of the discourse semantics relationships we have been looking at: theme versus rheme and contrastive versus non-contrastive, are signalled by pitch accent type, i.e. L+H* versus H*. However, these experiments give considerable support to our proposal in the last chapter that the theme/rheme relationship is conveyed by relative prominence. We have further seen that the prosodic relative prominence relationship is asymmetric, in line with our contention that the perception of prominence is mediated by its place in prosodic structure. Further, the results from Experiments 3 and 4 give clear evidence that relative prominence relationships hold over prosodic boundaries, supporting our arguments for recursive prosodic phrasing structure (see section 3.2.3). Does this mean then that we can reject the use of intonation contour *type* to signal information structure distinctions? On the evidence presented here, probably. However, we should return finally to the point made in the discussion of Experiment 2. It may be that the subtle $f0$ alignment cues found in Experiment 1, or other voice quality, etc cues not measured here, convey affective connotations correlated with themehood or Contrast. We will return to this point in Chapter 7.

4.3 Summary and General Discussion

The experiments reported in this chapter support the claim made in the previous chapter that theme/rheme status is indicated by relative prosodic prominence, not by pitch accent or boundary tone type. Experiment 1, a production experiment, showed several differences between the production of theme and rheme accents in nuclear position. Only productions with clearly identifiable $f0$ turning points were included. Rheme accents were higher than theme accents, although this result was only marginally significant. In addition, the L at the

[3] Except for rheme-theme order in the second clause. In this case the drop may have reached the 'floor' of the speaker's range.

start of theme accents was lower, and aligned later relative to the stressed syllable, than the L of rheme accents. There was a greater drop in $f0$ (T0-T1) with rheme accents than themes. Rheme accents were almost always followed by falling boundaries (L- or LL%), themes were equally likely to be followed by rising (H- or LH%) as falling boundaries. Experiment 2, a complementary perception experiment, showed that, of these differences, the only factor listeners could reliably use to judge the acceptability of a theme accent in nuclear position was peak height. Experiments 3 and 4 therefore pursued the nature of this relative height distinction. Experiment 3 reanalysed the materials from Experiment 1 to show that once 'weak' thematic accents were included, theme accents were clearly significantly lower than rheme accents. The weakly accented and deaccented cases were explained in terms of post-nuclear deaccenting, given that the rheme was in nuclear position. It was also found that peaks of Contrastive theme and rheme accents were correlated, suggesting this relationship could also be signalled by relative peak height. Finally, Experiment 4, a second production experiment, used new materials which allowed the direct comparison of paired and Contrastive themes and rhemes. Findings showed that, in all cases, the first accent in a phrase was higher than the second. However, in theme/rheme order the drop was less than in rheme/theme order. The asymmetry could also be seen in accents in a Contrastive relationship: there was no difference in accent height in theme/rheme order; while in rheme/theme order accent height dropped significantly. This asymmetry is consistent with our assumptions about the marking of relative prominence in relation to prosodic structure. Finally, the second production study also confirmed that rheme accents are marked by a greater fall after the accent, consistent with nuclear prominence marking.

We have shown how these findings accord with our claims in the last chapter about the mapping of the segmental string onto metrical prosodic structure. In particular, as set out in section 3.2.3, rhematic kontrasts try to align with nuclear prominence at the level of phrasing that includes both theme and rheme units. Prosodic prominence marking is asymmetric: among accents of equal, or near equal, height, the last will be perceived as nuclear. Post-nuclear prominences are deaccented (i.e. marked with little or no pitch movement) within the same phrase, and marked by much lower accents in a different phrase. This explanation neatly accounts for the production data from Experiments 1 and 3: in most cases the theme element was deaccented, as it occurred after its paired rheme. However, when it was accented, it was substantially lower. The importance of relative height as a perceptual cue was confirmed by the second experiment. Experiments 3 and 4 also showed that there was an asymmetric relative prominence relationship between Contrastive theme and rheme pairs. This relationship definitely held across prosodic phrase boundaries. These results, in particular the branching tree structures presented in Figure 4.10, therefore provide further evidence

for recursive prosodic phrasing structure, as argued for in section 3.1.2.

At first glance, these results would seem to hard to reconcile with the findings presented in section 4.1 about the marking of Contrast versus 'new information'. As discussed, (in studies which gave an acoustic analysis) Contrastiveness was associated with increasing the peak height of accents (Bartels & Kingston 1994, Watson et al. 2004, Krahmer & Swerts 2001, Rump & Collier 1996, Braun 2005). How does this fit with the finding in Experiments 1 and 3 that most themes, in Contrastive contexts, were either completely deaccented, or only weakly accented? In section 3.3.1, we discussed the need for a constraint-based model of prosodic structure formation. Contrastiveness evokes a *restricted* kontrast reading, which creates a pressure for accents to be realised in nuclear position, made emphatic or generally strengthened. Theme/rheme status, on the other hand, is signalled by a weak/strong prosodic relationship. Further, in post-nuclear position, there is pressure for elements to be deaccented completely. In the first set of materials, therefore, in most cases the constraint to deaccent thematic material in post-nuclear position was stronger than the constraint to emphasise Contrastive material. In the second set of materials, on the other hand, the two interacting contrasts were potentially more confusing. Further, in the rheme-theme case, the rhematic material was longer than in the first experiment, and the theme adverbial, so the theme was more likely to form its own phrase. Therefore, fewer post-nuclear themes were deaccented. This proposal is also consistent with our analysis of Hedberg & Sosa's (2001) findings regarding 'upstep' in section 4.1 above. If L+H* is used to mark themes which are raised for emphasis, then an emphasised rheme in the same phrase would be raised even further, hence the need for 'upstep' marking.

In Experiments 3 and 4, we found that the peaks of elements in a Contrastive relationship were correlated, and in Experiment 4, that this correlation seemed to exist in *addition* to the correlation between theme and rheme peaks found there. As noted by Umbach (2004), these elements are in fact linked in a particular sort of contrastive relationship, that of Correction. That is, there is a prosodic subordination relationship (which holds across phrases), between the *element being corrected* (weak), and the *correct element* (strong). This takes us back to our suggestion about the analysis of Steedman's (2006*b*) 'isolated themes' in section 2.3.3. There we argued that utterances such as *He's a good badminton player* might be not 'isolated themes', but one instance of a number of rhetorical relations between clauses, such as Nucleus-Evidence or Antithesis (cf. Mann & Thompson 1988). Possible support for this suggestion could be found here. We have shown that there are subordination relationships between clauses (such as Corrections) signalled by prosodic prominence *outside* of the theme/rheme division. That is, such relationships cannot all be reduced to instances of 'isolated themes' (although some may be best analysed this way). So here, within Steedman's

scheme, it is problematic to analyse *it isn't Henry Lombard* as an 'isolated theme' because this would not account for the large 'rhematic' accent on *isn't* (since Steedman assumes all accents are meaningful). As we have proposed at many points, there is a definite *correlation* with theme/rheme marking: *elements being corrected* tend to be thematic, and *correct elements* rhematic. However, this does not mean the two can be collapsed.[4] This difficulty for Steedman's theory is even more apparent with the sentences in Experiment 4.

Finally, our results support Ladd & Schepman's (2003) and Dilley's (2005) call to merge the ToBI categories L+H* and H*. As reported in section 4.1, their experiments have seriously questioned the phonetic basis for the distinction between these accents, i.e. the alignment and scaling of the preceding L. This still left to be explained the reported interpretative differences. As should be clear from the discussion above, these can be much more adequately accounted for by relative prominence within metrical prosodic structure. However, as discussed during this chapter, this does not rule out more subtle distinctions arising from illocutionary or affective connotations correlated with themehood or Contrastiveness (re section 3.3.2). We submit once more that these cannot be represented by ToBI pitch accents, but such effects may still exist. We will return to this in Chapter 7.

Our experiments here, then, leave us free to pursue more fully, and using much more diverse discourse material, the idea that relative prominence and phrasing, within metrical prosodic structure, are the primary prosodic cues to information structure in English. It is this that we will begin to do in the next chapter, where we describe the features of the data set used for the rest of our analysis, a subset of the Switchboard corpus.

[4]Note that, as discussed in section 3.2.3, it seems to be assumed without comment (by Steedman and others) that theme/rheme structure itself cannot be recursive, which might be able to account for this evidence.

Chapter 5

Switchboard in NXT: A Data Set for Model Development

In the preceding chapters, we have argued that prosodic structure is strongly constrained by information structure, but that it is also affected by low-level semantic and syntactic constraints, as well as constraints inherent on the prosodic structure itself (in particular see section 3.3.1). Therefore, in order to test our claims, we need to try to model these other constraints. This is not easy to do through controlled phonetic experiments. We need a large data set where many more of these interactions are exhibited. The *Switchboard* corpus is used because it is a large collection of spontaneous speech already annotated for a variety of discourse features. We have produced further annotation, of both semantic and prosodic structure, to adapt the corpus to our needs.

In Chapters 6 and 7, we use a data set derived from this corpus to test our predictions. Each data point is a word from a small portion of the corpus, with a large variety of discourse semantic, syntactic, prosodic and acoustic features. This chapter describes the features derived from existing annotations, together with our new prosodic and kontrast features, and acoustic features we extracted from the speech signals. We also discuss the prosody and kontrast annotation guidelines used.

The complete corpus can be seen as a new step in the field in terms of a collection of spontaneous conversations of such size and richness of annotation. It will be useful not only for the research questions being explored here, but for the analysis of other diverse linguistic phenomena.

5.1 The *Switchboard* Corpus in NXT

The Switchboard Corpus (Godfrey et al. 1992) consists of spontaneous telephone conversations between American English speakers, distributed as stereo speech signals with an orthographic transcription per channel. Each conversation was between two previously unacquainted paid participants on a topic chosen from a pre-determined list.

A subset of 642 conversations, just over 830,000 words, was annotated for part-of-speech information, syntax structure and disfluencies as part of the Penn Treebank Project (Marcus, Santorini & Marcinkiewicz 1993, Taylor, Marcus & Santorini 2003). This subset has been converted into Nite XML Technology format (NXT) (see Carletta et al. 2004). NXT provides an integrated data representation, along with tools for querying and extracting data from the corpus (Carletta, Evert, Heid, Kilgour, Robertson & Voormann 2003, Carletta, Evert, Heid & Kilgour in press). The subset has also been annotated for dialog acts (Shriberg et al. 1998), and a smaller portion for information status (Nissim et al. 2004); both of which are now in NXT format.[1] The information status subset was also used for our kontrast annotations (see section 5.6).

In addition to these text-based features are annotations derived from the acoustic signals. These include a corrected transcript time-aligned at the word level (the MS-State transcript, Deshmukh, Ganapathiraju, Gleeson, Hamaker & Picone 1998, Harkins 2003); automatic phone and syllable alignments produced using the Sonic speech recognition system (Pellom 2001);[2] and prosody annotation (Ostendorf, Shafran, Shattuck-Hufnagel, Carmichael & Byrne 2001). All of these are now in NXT format. In order to integrate both sets of annotations, the Treebank and MS-State transcriptions had to be aligned. Because of differences in the way they were produced, this alignment was not perfect, so unaligned words were lost (2%). We updated Ostendorf et al.'s (2001) existing prosody annotations using standards more suitable for our research questions, and included more conversations of our own, as detailed in section 5.4.

An overview of all the layers of annotation and their relationship in the NXT corpus is shown in Figure 5.1. More details about the existing annotations can be found in Appendix C.

5.2 Discourse Semantic Features

A variety of discourse semantic features were extracted from the existing layers of corpus annotation using NXT queries. Using the *disfluency* codings, words were categorise

[1] Many thanks to Neil Mayo, Jean Carletta, Shipra Dingare and Colin Matheson who variously carried out the conversions to NXT format described here.

[2] Many thanks to Jason Brenier for producing these alignments.

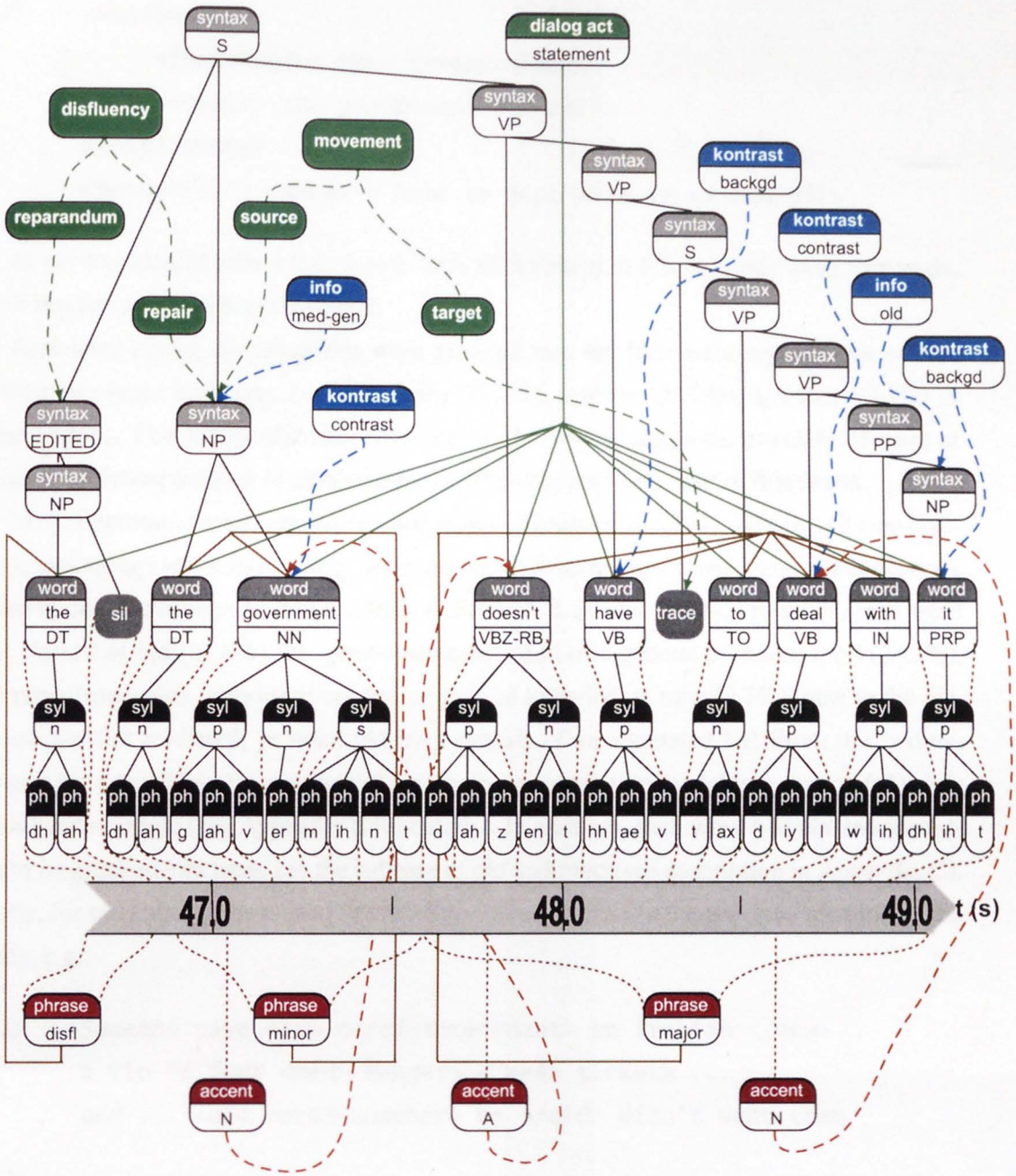


Figure 5.1: Overview of annotations from a small sample of the *Switchboard* corpus as represented in the NXT data model (timing not exact). (Coreference and trigger relationships not shown, however, these work similarly to disfluency and movement relationships).

as reparanda, repairs or notdisfl, e.g.:

(5.1)
```
<disfluency>
    <reparandum> the- </reparandum>
    <repair> the government </repair>
</disfluency>
<notdisfl> ...doesn't have to deal with it </notdisfl>
```

In our experiment subset (i.e. words with all features), 0.6% of words were reparanda, 6.1% repairs, and 93.4% not disfluent.

Annotated *dialog act* categories were grouped into the four main types in our subset, as the other types were rare, i.e. statement (70.1%), opinion (22.2%), question (3.1%) or other (4.6%). This feature did not prove very useful for our purposes, probably because of the overwhelming number of statements, and the very small number of questions.

More important for us was information status (Nissim et al. 2004). NPs in 147 conversations, averaging eight minutes long, were annotated from the text alone, using classifications based on the taxonomies of Prince (1992) and Eckert & Strube (2001). From this, each word was classed as old (21.6%), i.e. previously mentioned or a generic pronoun; med (24.2%), inferable from other introduced entities or general knowledge; new (9.7%), new to the discourse and not mediated; or noinf (44.6%), outside of an annotated NP. From this coding, we also introduced two features useful for the phrase prediction model (see section 6.1): *info bound* and *next info* (see Appendix E for details). In addition, each word was classed according to its grouped info type, i.e. the subtype of old and med (see description in Appendix C). Lastly, for old entities, the coding included *coreference* links to the previous mention of the entity, e.g.:

(5.2)
```
Someone gave <new coref=antecedent> an Iranian </new>
a tip of four <med> Rangers </med> tickets ...
and ... <old coref=anaphor> he </old> didn't want them
```

From this we derived a *dist coref* feature, the number of words since the last mention. This did not prove useful, however, probably because of the small proportion of words marked for this feature. Along with these disfluency, dialog act and information status features, the main class of discourse semantic features used were *kontrast* features, described in section 5.6.

5.3 Syntactic Features

Syntactic features were derived from the Penn Treebank coding. Each word was classed according to its part-of-speech group (*POS gp*), using an NXT query. These were grouped into seven major types in order to improve their statistical power: NN (noun) 28.3%, VB (verb) 25.9%, PR (pronoun) 17.9%, DT (determiner) 6.5%, JJ (adjective) 8.2%, RB (adverb) 13.0% and XX (other) 0.1%.

From the syntax coding, we classified each word according to its broad clause and constituent type. To do this, we used the coding in TGrep2 format (Rohde 2005).[3] TGrep2 offers a faster way to extract parse trees matching specified patterns, since it assumes a non-crossing tree structure, and is therefore more computationally efficient than NXT. The method used was to build patterns which identified words in each of the different clause and constituent types, until all the words in a conversation were classified, and then to test these patterns on the next conversation and repeat the process. Because certain parse tree patterns are very rare, some unidentified words were lost (1.3%). This deterministic method may have introduced inaccuracies. However, it was felt to be sufficient since our primary focus was not syntactic parsing, and we did not have to commit to the theoretical basis of one or other statistical parser.

Clause types were those noted in the literature as having distinctive prosodic characteristics (see Chapter 2), defined as follows (see Appendix C for a description of Penn Treebank tags):

1. **Parenthetical:** (0.5%) in a PRN clause.
2. **Adverbial:** (18.7%) in a clause dominated by SBAR-ADV, -TMP, -LOC or -PRP; a PP; S-ADV; ADVP or WHADVP.
3. **Relative:** (4.5%) in a clause with a non-empty subject, which is dominated by an NP.
4. **Complement:** (21.6%) in a clause dominated by a predicate (VP or ADJP-PRD) (including *that*-clauses, *for*-clauses and infinitival clauses).
5. **Main:** (54.4%) in a clause dominated only by the main S or SQ in the sentence, and not by any of the clause types above.

Some words were included in more than one clause type, in which case they were classed in the order above, e.g. if a word was both comp and main, it would be classified as comp.

Constituent type encoded the main relationships between elements in a clause described in standard syntactic structure theory:

[3] Many thanks to Neil Mayo and Jean Carletta for doing the translation.

1. **Subject:** (18.7%) in a NP-SBJ, immediately dominated by an S.
2. **Predicate:** (34.4%) in a constituent dominated by a predicate (VP or ADJP-PRD), which is immediately dominated by an S, with no NP or PP nodes between the word and the predicate. Includes any modals or adverbs immediately before the predicate.
3. **Object:** (24.0%) in an NP (without adjunct function tags) immediately dominated by a VP, which is dominated by an S.
4. **Adjunct:** (22.1%) in an PP or NP-DIR, -LOC, -MNR, -PRP, or -TMP immediately dominated by a predicate, an S or an ADVP.

In addition to constituent type, separate TGrep2 patterns were used to identify the *head* of each constituent, i.e. the last word immediately dominated by the XP dominating the whole constituent with same type as that XP, e.g. if the whole phrase was an NP, this would be the last N immediately dominated by that NP. Heads are often claimed to be less accentable than their arguments. As well as clause and constituent type, features were extracted showing the position of the word in relation to its clause and constituent (see Appendix E).

5.4 Prosody Annotation

To make the corpus useful for us, we needed annotation of basic prosodic features. Unfortunately, both the development of such guidelines and the annotation itself are very time consuming, so only a small portion of the corpus has been completed. It was felt that automatic prosodic event detection algorithms are not yet mature enough to be used instead, at least for spontaneous speech. This placed the main limit on our experiment data set, which only included the 18 conversations annotated for prosody. Below we describe our guidelines, and the annotations, most of which were adapted from those previously done (Ostendorf et al. 2001). Firstly we briefly review related projects annotating spontaneous speech in English.

5.4.1 Related Work

As reported in section 3.1.3, annotation of pitch accents and phrase boundaries in English using ToBI is now well-established, with agreement ranging from 80-93% across a variety of speakers and corpora. However, most studies use read or highly structured speech, or only a small proportion of spontaneous speech. There have been relatively few attempts to annotate fully spontaneous speech, and reported agreement is more variable. As was noted above, it is not standard within this system to mark the nuclear accent. We are aware of one earlier study

which did try to annotate nuclear accents (Brown et al. 1980). They did not report annotator agreement directly, although they did include extensive discussion of agreement problems which we refer to below.

At least two previous studies labelled a small subset of isolated utterances from Switchboard, just over an hour of speech (the ICSI corpus, Greenberg, Hollenback & Ellis 1996). We did not use this subset as its fragmentary nature was not good for dialogue analysis, however, their results are useful for comparison. Taylor (2000) labelled a number of corpora, including this subset, for pitch accent and boundary location. He assessed agreement using the DCIEM Maptask (Bard, Sotillo, Anderson, Thompson & Taylor 1996), which comprised spontaneous, direction-giving dialogues. He reports agreement of 82% correct with 58% accuracy for pitch accents and 83% correct with 64% accuracy for boundary tones. This is lower than earlier studies, but he suggests the biggest difficulty was 'level accents', i.e. clear prominence with little discernible pitch movement. With boundaries, the biggest difficulty was the classification of disfluent or abandoned phrases. More recently, Yoon, Chavarría, Cole & Hasegawa-Johnson (2004) annotated this same subset with a modified version of ToBI, i.e. pitch accents, phrase accents and boundary tones were marked as unitonal H or L. They report agreement of 89% on presence of a pitch accent, and 87% on pitch accent type. For phrase accents agreement is 86%, and 89% for presence and type of boundary tones. Note, however, that pairwise agreement is an easier and less reliable measure than kappa, which we describe when reporting our own agreement figures below (for problems with other measures, see Carletta 1996).

The only project we are aware of that has annotated whole Switchboard conversations is Ostendorf et al. (2001), who annotated 59 dialogues with simplified ToBI labels. Annotators labelled a break index tier, identifying 0, 1, 1p, 2, 2p, 3 and 4 breaks; and a tone tier, labelling L-, H- phrase accents at 3 breaks as well as L% and H% boundary tones at 4 breaks. At 3 breaks they could also use !H- phrase accents for a mid-range pitch fall after a high accent. Accents were identified using a *, or *? for a weak accent. Tonal pitch accent type was not labelled. Since there was only one annotator, no agreement figures are available. The researchers involved kindly agreed to include these annotations in the present project. So, for the 12 conversations which overlap with our kontrast set, our annotators adapted these existing prosody annotations using the guidelines below which were more specifically attuned to the needs of our project.

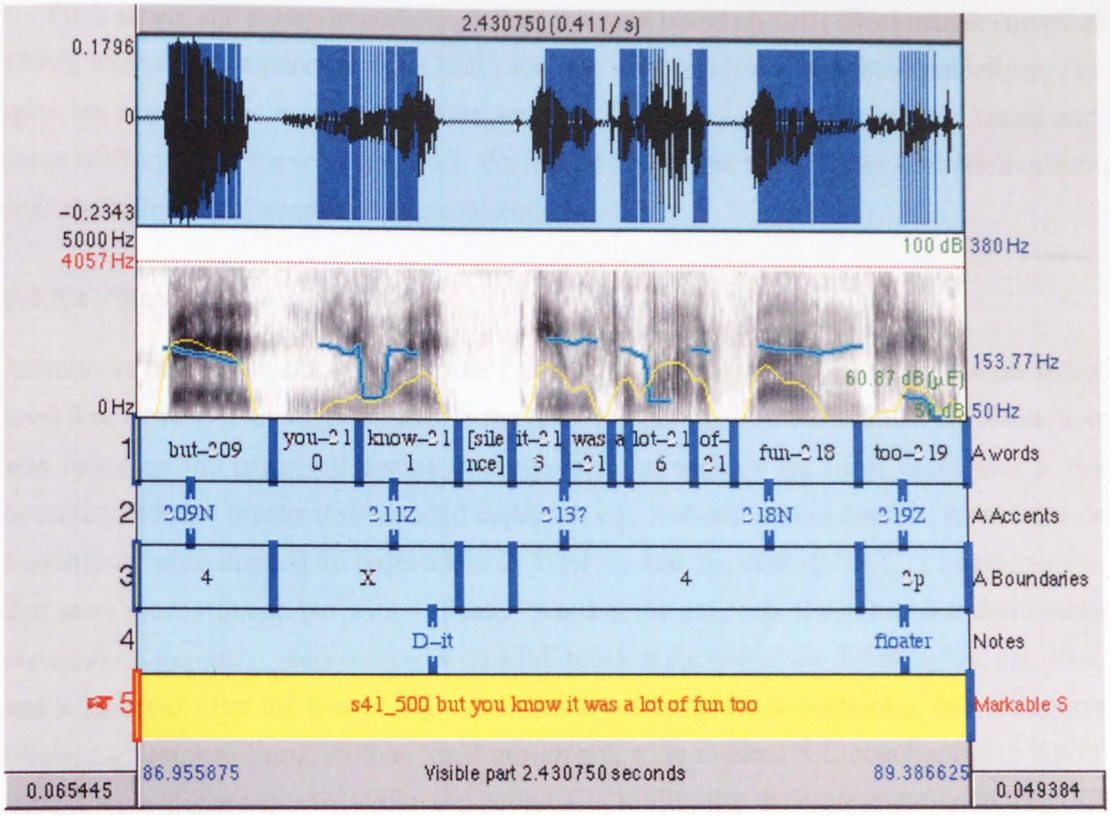


Figure 5.2: Screenshot of the Praat labelling tool for original prosody annotations. Annotators were given the Words and Markable Sentences tiers, and had to mark indexed accent type in the Accent tier, and phrase boundaries and type in the Boundaries tier, as well as optional notes in the Notes tier. The screens at the top show the waveform, $f0$ trace (blue line), intensity curve (yellow line) and spectrogram (grey shading).

5.4.2 Annotation Scheme

Our annotation scheme was developed in response to the research questions raised in Chapter 3.[4] Annotations were carried out for each speaker separately on the MS-State transcript using *Praat* (Boersma & Weenink 2003), and then later converted to NXT format.[5] The *Praat* tool allowed visual presentation of acoustic information including the pitch track and intensity curve (see Figure 5.2). To save time, in the conversations that were not converted from Ostendorf et al.'s (2001) set, annotators marked only those sentences which containing words marked for kontrast (see section 5.6.2.1).

[4]Please note that this scheme was developed, and the annotations carried out, jointly by the author and Jason Brenier.

[5]Many thanks again to Neil Mayo and Jean Carletta for performing this conversion.

Pitch accent and phrase boundaries definitions were based on ToBI (Beckman & Hirschberg 1999), like with Ostendorf et al.'s (2001) scheme. We decided to use these guidelines, despite the reservations expressed in Chapter 3, as they are well established and tested and come with a pool of trained annotators. We felt they would be adequate for our requirements with the following changes and augmentations.

5.4.2.1 Boundaries

Annotators marked breaks as one of four types. For fluent phrases, they marked break index level 3 or 4 (minor and major phrases in the NXT representation). As in ToBI, the distinction was based on the perceived degree of disjuncture, as well as the tonal movement at the boundary. Phrase breaks that sounded disfluent, e.g. cut-offs before restarts, repetitions or hesitations, were marked 2p (equivalent to ToBI 1p and 2p, disfl in NXT). Longer phrases that were mostly fluent, but with a 'floater' word at the end, e.g. *the plants just don't make the winter time, and...*, were marked with a full break at the end of the fluent region, e.g. *time*, and a 2p break after the floater, e.g. *and*. Finally, short phrases containing only discourse fillers, e.g. *um*, *you know*, with no tonal movement, were marked X (backchannel in NXT). Breaks were aligned exactly with word boundaries to simplify the representation in the NXT (see Figure 5.1).

5.4.2.2 Accents

The presence or absence of accents, and not tonal pitch accent type, was marked. Annotators labelled each accent with an index, so it could be unambiguously associated with that word they heard it on (see Figure 5.2).

Unlike in ToBI, annotators marked one accent in each phrase as being nuclear (N), defined as the *structurally*, not *phonetically*, most prominent (cf. Brown et al. (1980) and discussion below). They were told to listen for the accent that sounded the most important, normally the right-most one. After some discussion and practice, annotators were able to use this concept effectively. However, certain cases proved problematic: particularly when there was an early emphatic high accent and a later, downstepped nuclear accent (see Figure 5.3). We therefore introduced the pre-nuclear label (PN) to mark the earlier accent, which annotators found helpful.

Annotators also had some difficulty with weak pre-nuclear accents, where there was definitely some prominence, but little pitch movement, or vice versa. Such non-nuclear accents were marked as possible accents (Q). Finally, although in general each phrase had at least one (nuclear) accent; disfluent phrases, or phrases that only contained filled pauses (e.g. *um*, *er*), where no words sounded accented, were marked with a Z (see Figure 5.2), i.e. unaccented.

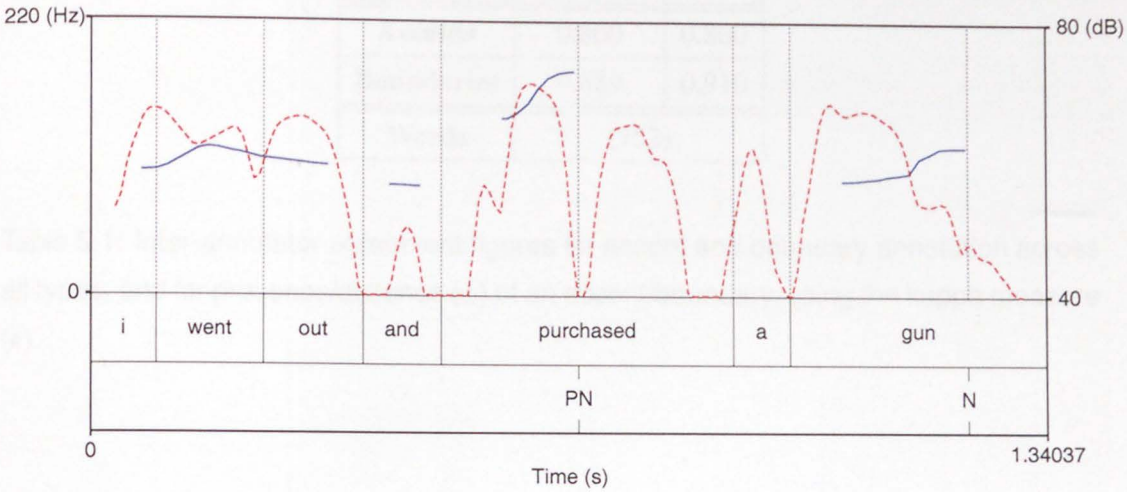


Figure 5.3: Example of distinction between PN and N accents (blue line is the $f0$ trace, the dashed red line the intensity curve).

5.4.3 Annotation Process and Annotator Agreement

Most of the annotations were done by a paid post-graduate linguistics student at the University of Edinburgh, with experience using the ToBI guidelines, and a smaller number by the author.[6] After an initial training and discussion period, annotations took 8-10 hours for original conversations, and 6-8 hours for converted conversations.

One original conversation side was used to check annotator agreement, see Table 5.1. The kappa statistic (κ) is reported because it is more reliable and easily comparable than others commonly used (see discussion in Carletta 1996). Kappa measures pairwise agreement among annotators, correcting for expected chance agreement:

$$\kappa = \frac{P(A) - P(E)}{1 - P(E)} \tag{5.3}$$

where $P(A)$ is the proportion of times the annotators agree, and $P(E)$ the proportion of times we would expect them to agree by chance. $\kappa > 0.8$ is said to show "good reliability", while $0.67 < \kappa < 0.8$ "allows tentative conclusions to be drawn" (Carletta 1996). We can see that agreement on breaks was better than for pitch accents, but both are good. It is commensurate with the figures reported above for related studies, showing our guidelines were successful. There is little difference in κ for all types versus binary presence/absence

[6]The first sample conversation side was initially done by Jason Brenier, then later checked and amended by the author.

	All Types	±
Accents	0.800	0.800
Boundaries	0.889	0.910
Words	(752)	

Table 5.1: Inter-annotator agreement figures for accent and boundary annotation across all types, and for presence/absence (±) of an accent/boundary, using the kappa measure (κ).

of both accents and boundaries, showing successful discrimination among types.

Like Taylor (2000), annotators found the most difficulty in classifying 'weak' accents. The 'Q' label improved consistency, however, the difficulty with such cases is really inherent in trying to label 'accents'. As discussed in section 3.1.2, pitch movement is one way to mark phrasal prominence, not equivalent to it, which ToBI can tend to imply. Alternative schemes such as RaP (Dilley & Brown 2005), which explicitly separate the labelling of rhythmical prominence and intonational movement, could be a way forward.

In section 3.1.2, we also showed that the structurally most prominent syllable can be distinct from the acoustically most prominent. However, as mentioned above, this distinction is hard to make with particularly strong pre-nuclear accents. The 'PN' label helped, but there remained difficult cases, e.g. in Figure 5.4, *kind* is definitely more acoustically prominent than *you*. In such cases the annotator had to turn to the semantics. Here, the speaker was more interested in *you* than in *kinds of things*. Therefore, *you* as nuclear. This may seem circular. However, it is difficult to see how it could be avoided since interpretation and nuclear accent placement go hand-in-hand. In fact, Brown et al. (1980), on the basis of their prosody annotation experience, conclude that tonics cannot be reliably identified in cases where phonetic and semantic cues conflict. They note, however, that in their dialogues the 'new' information (roughly equivalent to foci) tended to come at the beginning of utterances, i.e. against the usually noted trend. It may be that this is a peculiarity of the genre they were using (responses to questions in an interview) or the dialect (Scottish English). This does not seem to be the case with our data, and it may be that it confused the expectations of annotators. On the other hand, if this were the case, the reversal of the 'default' nuclear accent position could be accommodated within our framework. That is, we would assume the right-branching 'bias' is learned from the usual frequency distribution, which may be reversed in certain contexts.

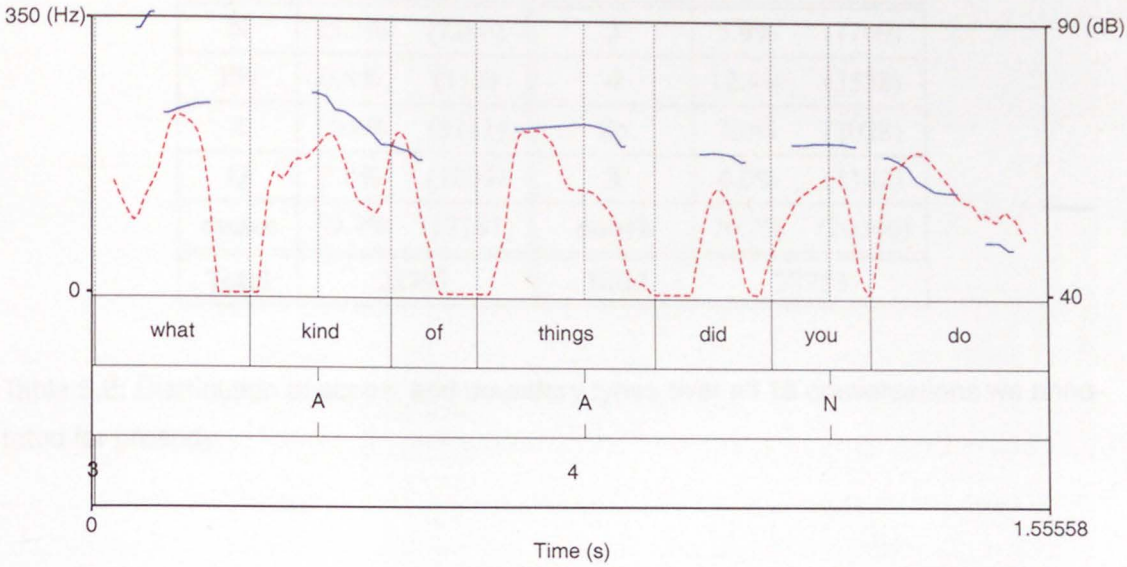


Figure 5.4: Ambiguity in the placement of the nuclear accent in cases with a strong pre-nuclear accent, i.e. between *kind* and *you* (blue line is $f0$, dashed red line intensity).

Finally, a shortcoming of our scheme is that we did not mark the perception of nuclear prominence over a larger recursive phrasing structure, which is important to our claims. Although we did mark two levels of break strength (3 and 4), initial analysis showed this could not be used to reliably identify larger phrasal structures. Unfortunately, it is hard to see how such structures could be consistently and efficiently classified over large amounts of data; especially since the perception of such structures is influenced by both acoustic and semantic features. Brown et al. (1980) further note that annotators varied a lot in nuclear accent placement in long utterances. We would claim this difficulty was caused by ambiguity of phrasing, not nuclear accent placement. Resolution of this problem is beyond the scope of this work.

5.4.4 Distribution of Prosodic Types

Table 5.2 shows the frequencies of accent and boundary types in the set of conversations we annotated for prosody (there are slightly more accents than boundaries because some words had two accents).

Accent	Freq	Boundary	Freq
N	25.3% (7296)	3	5.9% (1709)
PN	0.4% (115)	4	12.4% (3558)
A	10.8% (3111)	2p	7.0% (2028)
Q	3.8% (1094)	X	4.0% (1142)
unacc	59.7% (17181)	nobrk	70.7% (20346)
Total	28797	Total	28783

Table 5.2: Distribution of accent and boundary types over all 18 conversations we annotated for prosody.

5.5 Prosodic and Acoustic Features

The prosodic annotation, along with the automatic phone and syllable alignments (see details in Appendix C), was used to extract a variety of prosodic and acoustic features (see Appendix E for full list). Firstly, all words in phrases labelled 2p or X were excluded from the data, so the final data set only included words in fluent phrases. This was because, as noted in Chapter 3, our theory decribes fluent speech, we have no particular predictions about how this interacts with disfluency.

Using the phrase break annotation, we extracted a number of features describing each word in relation to the phrase it was in; as well as the same set of features for the first syllable and phone in each word, and the duration of the phrase. We measured the *speech rate* as the total number of syllables relative to the duration of the phrase. We also extracted a number of features that tried to capture the eurhythmic properties of the word in relation to the surrounding prosodic structure, i.e. the constraint that each phrase should have approximately the same number of syllables.

Using the accent annotation, we extracted *accent group* (accent, nuclear, no accent) and *accent status* (pre-, post-, nuclear) features. In exploratory models, we tried grouping weak accents (Q) with accents and unaccented words, and concluded that they seemed to pattern more strongly with accents. Therefore, in all reported models, accents include Q accents. We further decided to include PN accents with nuclear accents, since they seemed to pattern in the same way. This concurs with our predictions. However, the status of PN accents does still need to be worked out (see further in section 7.2). We also included features meant to capture the rhythmical properties of the word in relation to the phrase.

Word and phrase level acoustic features were extracted automatically using Praat scripts (Boersma & Weenink 2003). Data was lost when the word was too short for Praat to extract $f0$. To alleviate this, words less that 300ms long were given a window of 50ms on each end. All pitch values were normalised as a percentage of the speaker's logged range, to account for inherent speaker differences (see Ladd 1996, ch. 7). This was calculated by extracting all $f0$ values (10ms intervals) for each speaker in all their conversations in the corpus (around five). These values were then logged and ordered. Values more than 2.5 std. dev. away from the mean for that speaker were excluded to remove outliers. Ordered values were then grouped in 1000 equal sized bins, to be used as 'look-up' tables for the logged pitch values extracted for each word and phrase; e.g. if the extracted value fell in the 456th bin, the normalised value would be 0.456. Values which did not fall within the range of these bins were excluded as probable pitch errors (around 12% total not found or excluded). This method was used as pitch values are known not to be normally distributed, so Z-scores were not appropriate. Here the actual distribution of pitch values is directly reflected in the normalised measure. Logged values were used because this is closer to human perception of pitch than a linear range (see Ladd 1996, ch. 7). Intensity values were normalised by dividing by the mean of the mean intensity of each of that speaker's words in the conversation. This was done rather than extracting all intensity values, because then the mean would be swamped by the low (but still measurable) levels of intensity when the speaker was not speaking. Word duration was measured relative to the number of syllables in the word. Although simple, this normalisation method seemed to be effective. Lastly, all acoustic measures were multiplied by ten, to make them more interpretable in comparison to categorical variables in the regression models.

The pitch at the annotated accent peak was extracted, and normalised as above. The position of the peak (H) was normalised relative to the stressed syllable as follows:

$$naccH = \frac{H - C0}{C1 - C0} \tag{5.4}$$

where $C0$ is the beginning of the stressed syllable, and $C1$ the end of the stressed syllable. An approximate measure of the accent L(ow) position was also included. This was calculated by normalising the time of the (automatically derived) pitch minimum in the word using the same procedure, where the L occurred before the H. In addition, positional features were only included where the maximum pitch for the word was in the top 50% of the speaker's range; since annotators were instructed to mark 'flat accents' in the middle of the stressed syllable, other positional measures would not be accurate. Obviously this resulted in a large loss of data, so negative results with these features with less persuasive than with other features.

5.6 Kontrast Annotation

A central concern in this thesis is describing the relationship between focus and prominence. In section 2.2.1.3, we showed how focus-related phenomena, including wh-focus, the scope of focus-sensitive operators and relative givenness, can be explained by Alternative Semantics (Rooth 1992). In each case, the focus is a *kontrast*, i.e. it introduces a presupposition of alternatives to the kontrasted phrase in the discourse context. Our annotation scheme begins with the assumptions of Alternative Semantics. The broad plan was to encode whether each markable element was kontrastive or not. However, the annotators' specific task was to identify instances of each discourse phenomena, not kontrast per se, allowing for potential differences in their behaviour. Below we describe the annotation scheme, including the selection of markable elements. We then report on its success in terms of annotator agreement and the experience of the annotators. Finally, we briefly describe the distribution of the annotations. Before this, we report other efforts to annotate focus-related phenomena.

5.6.1 Related Work

To our knowledge there have been few attempts to annotate information structure in unrestricted discourse. Our approach is therefore novel, stemming from the theoretical work in section 2.2.1.3. However, here we briefly review related efforts.[7]

Grosz and Hirschberg (Grosz & Hirschberg 1992, Hirschberg & Grosz 1992, Nakatani, Hirschberg & Grosz 1995) adapted the discourse structure theory of Grosz & Sidner (1986) (see section 2.2.3.1) to annotate a variety of spoken corpora, including some spontaneous speech. They segmented discourses according to their *linguistic structure*, analogous to *theme/rheme* pairs at the local level (see section 5.7). In later work (see also Hirschberg 1993, Nakatani 1994), they use this structure to identify *contrasts*, i.e. NPs which are *new* at the local level, but *given* on the global level, re Grosz & Sidner's (1986) *attentional structure*. There was a high correlation between these contrasts and pitch accenting. They report no significant difference between annotations of different labellers. For our purposes, we felt their definition of *contrast* would only capture some thematic kontrasts on NPs, and therefore miss phenomena we were interested in.

Hedberg & Sosa's (2001) scheme is closer to ours (see section 4.1.2). They appear to annotate units of arbitrary length from words to sentences. Each unit is annotated for topic/focus, and ratified/unratified/contrastive. Ratified/unratified is covered by *information status*. Though not defined, *contrastive* seems to apply to explicitly contrasting units in the

[7] Note this does not include *information status* annotation, i.e. *given*, *new* or *inferable* relative to the discourse. For a review of this literature see Nissim et al. (2004).

context, similar to our *contrast*. Their topic/focus is close to our theme/rheme (see section 5.7). Since the single annotator was one of the authors, agreement figures are not given.

Recently, Zhang, Hasegawa-Johnson & Levinson (2006) annotated a small spoken corpus for *focus-kernels* and *contrast*. The corpus consisted of 'wizard-of-oz' dialogues between children aged 9-12 and a simulated talking head. Focus kernels were content words "that contain information not already available in presupposition, if any, nor in the preceding words of the utterance", analogous to infostatus *new*. *Contrast* pairs were words which were semantically independent, and with a 'common integretor', similar to our *contrast*. They also required that the pair be syntactically parallel, e.g. "The large$_C$ gear spins left, and the medium$_C$ gear spins right", which we did not think was justifiable for our scheme. They do not report annotator agreement for contrast.

The annotation of information structure on written corpora in Czech as part of the Prague Dependency Treebank (Böhmová, Hajič, Hajičová & Hladká 2001, Buráňová, Hajičová & Sgall 2000), has the most similar theoretical assumptions to ours. They annotate topic-focus articulation on lower layers of morphemic, syntactic and dependency relationships between syntactic nodes, elegantly pre-determining the scope of contrast. Each node is marked as contextually bound or unbound (theme/rheme). Each bound node is then marked as contrastive or non-contrastive, i.e. explicitly contrasting in the context. Good agreement is reported on node annotation, 80-90% accuracy (Veselá, Havelka & Hajičová 2004). However, agreement on the trees that determine these nodes is only about 30%. In other words, it is much easier to determine bound/contrast status than to determine the scope of the different units. The scheme only identifies contrast in themes, while we were also interested in contrast in rhemes. It can also only identify discourse phenomena associated with contrastive topics, e.g. focus-sensitive adverb association, after annotation (see discussion in Hajičová & Sgall 2004), making it more theory dependent than ours.

As can be seen, previous schemes have annotated topic, focus and contrast. These cover some, but not all, of the phenomena discussed in the literature on kontrast. There seems to be no general agreement on the unit of annotation, with different schemes proposing the word, the NP, a pre-determined syntactic node, or units of variable length at the discretion of the annotator.

5.6.2 Annotation Scheme

Annotations were carried out on the subset of Switchboard used in *information status* annotations, using a scheme loosely based on the one used there (Nissim 2003). Words to be annotated (markables) were selected and displayed in an NXT tool.[8] This allowed annotators

[8]Thanks to Jonathan Kilgour for developing the tool.

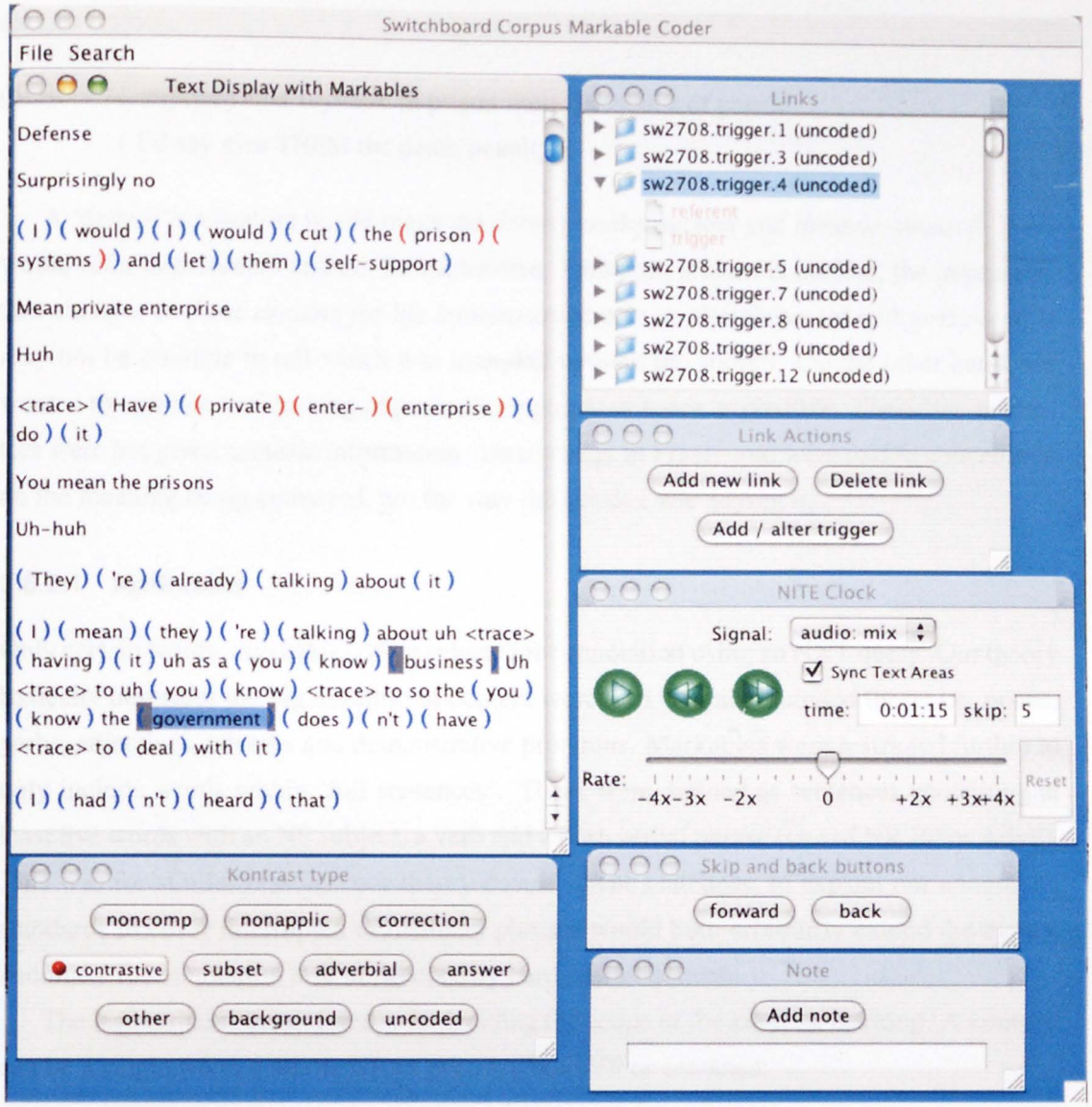


Figure 5.5: Screenshot of the NXT kontrast annotation tool, showing markable words and NPs in brackets on the left, buttons to label kontrast type at the bottom, buttons to mark trigger links top right and the embedded playback tool bottom right.

to mark kontrast type and trigger links while listening to the conversation (see Figure 5.5). It then automatically produced NXT-conformant files from these annotations.

It was decided that it would be too difficult for annotators to use text alone, rather than text and speech, given the highly ambiguous nature of speech, particularly spontaneous discourse. Further, there was the danger of bias towards 'default' prosody; for instance here, where the speaker is discussing the merits of life imprisonment versus the death penalty for

serious crimes:

(5.5) ...anybody that says life in prison with no chance of parole
(I'd say give THEM the death penalty)

A 'default' realisation would make *the death penalty* nuclear and *them* de-stressed. *them* would refer to *prisoners* and not be kontrastive. However, if *them* is nuclear, the interpretation changes so *those arguing for life imprisonment* is in an alternative set with *prisoners*. It may not be possible to tell which was intended without the speech. On the other hand, we wanted to separate the marking of prosodic emphasis as much as possible. Therefore annotators were not given acoustic information visually (e.g. in Praat); and were told to concentrate on the meaning being conveyed, not the way the speaker was saying it.

5.6.2.1 Markables

Only certain words, *markables*, were selected for annotation using an NXT query. Our theory basically describes the organisation of content words, so we only included these; i.e. nouns, verbs, adjectives, adverbs and demonstrative pronouns. Markables were restricted further to only include words within 'full sentences'. These were defined as sentences containing at least five words with an NP subject, a verb and a post-verbal phrase (one of NP, PP or AdvP). This was for similar reasons: our theory covers fluent sentences, to expand our annotation standards to cover interrupted or disfluent phrases would both arbitrarily extend the theory, and make the annotation task unnecessarily hard for annotators.

The last, rather difficult issue, was deciding the scope of the kontrast marking. A kontrast can be a single word, a whole NP, or even a whole VP or sentence:

(5.6) (TWO thousand and ONE was a good *MOVIE*) (IF you had read the **BOOK**)

(5.7) they make it really easy for people to uh...
(to get CREDIT CARDS) (*especially* **college students**)

(5.8) (They initially *START out in the BARNYARD kind of setting*) ...
(and they **wind UP**) (**all over CREATION**)

In (5.6), the contrast is between *movie* and *book*, not between *good movie* and *book* (i.e. there is no implication the book is bad). Therefore the kontrast acts at the word level. In (5.7), however, the focus of *especially* is the whole NP *college students*, i.e. there is no implication *college students* are targeted over other *students*, or *students* over others in *colleges*. In (5.8), the comparison is between the two VPs *starting out in the barnyard* and *winding up all over creation*.

It was decided it would be too difficult to maintain consistency if the scope of kontrast marking were unrestricted. Therefore, we went for a "half-and-half" solution. Markable elements were shown at both the word and NP level (see Figure 5.5), and annotators marked *either* the word or the NP, depending on which they thought more naturally applied. Where the unit was larger than the NP, they marked the word or words the speaker emphasised the most, e.g. *start*, *barnyard*, *up* and *creation* in (5.8).

5.6.2.2 Exclusions

Other words not covered by our theory were marked non-applicable, and excluded from further analysis. These included *false starts*, *hesitations*, and *idiomatic phrases*, e.g. *in fact* or *you know*. Annotators marked the latter sparingly, only for highly formulaic usages where the words had very little relation to the meaning.

5.6.2.3 Trigger Links

In all categories except for answer, other and background, annotators marked a trigger link between the word or NP that motivated the category assignment (the trigger) and the element being marked (the referent). For example, the trigger link in (5.6) would be:

(5.9)
```
<trigger>
    <trigger -> movie>
    <referent -> book>
</trigger>
```

This helped motivate the category assignment. We also thought the prosodic realisation of a referent might change depending how far away the trigger was.

5.6.2.4 Categories

Annotators were asked to identify words or NPs which were "salient with an implication that this salience is in comparison or contrast to other related words or NPs explicitly or implicitly evoked in the context". These words or NPs were marked with one of the following kontrast types, all other words were background. The types were derived from the literature in section 2.2, particularly Rooth (1992).[9] In addition to the guidelines, there was a decision tree for difficult cases; especially where more than one category seemed to apply (see Appendix D). This ranked the types based on their perceived relative salience.

[9] Rooth did not include *subsets*, however it was felt that some *contrast* examples could more naturally be classified as *subsets*.

5.6.2.4.1 correction The speaker's intent was to correct or clarify another word or NP just used by them or the other speaker. Corrections are an exaggerated or extended form of contrast, with potential realisation differences (see discussion in Umbach 2004). Annotators marked the correction and then the word or NP being corrected as the trigger. For example, in (5.10) the speaker wishes to clarify whether her interlocutor really meant "hyacinths" as opposed to any other bulbs.

(5.10) (A)... it was a *hyacinth* have you ever seen those? Oh they are pretty in the Spring but the leaves I do not like them...
(B) (now are you sure they're **HYACINTHS**) (because that is a BULB)

5.6.2.4.2 contrastive The intent was to contrast the word with a previous one which was (a) a current topic, and (b) semantically related to the contrastive word, such that they both belonged to a plausible set. This could be fairly abstract if the intended contrast was clear. Contrasts could be realised in two ways. In cases like (5.11) the speaker highlights both the trigger and the referent in making the contrast, so both words were marked as contrastive, and a trigger link created between them.

(5.11) (I have got SOME in the *BACKYARD* that) (bloomed BLUE) (Which I WOULD have liked those in the **FRONT**) (because they match my PORCH)

In cases like (5.12), the contrast only works backward. B contrasts recycling in her town "San Antonio", with A's town "Garland", from the set *places where the speakers live*. So *San Antonio* was marked as contrastive and then linked to the trigger *Garland* (which was not contrastive).

(5.12) (A) I live in *Garland*, and we're just beginning to build a real big recycling center that recycles everything imaginable...
(B) (YEAH there's been) (NO emphasis on recycling at ALL) (in **San ANTONIO**)

5.6.2.4.3 subset The word was (a) a current topic, and (b) a member of a more general set mentioned in the context. Again, the set could be fairly abstract if the intended set-subset relationship was clear. In (5.13), the speaker introduces the general set "three day cares", and then gives a fact about each. *Two in Lewisville*, *one in Irving* and *the second one* are all subsets.

(5.13) (THIS woman owns *THREE day cares*) (**TWO** in Lewisville) (and **ONE** in Irving) (and she had to open **the SECOND one** up) in Lewisville (because her WAITING list was) just like you like (a YEAR old)

As in the contrastive case, if both the trigger and the referent were said by the same speaker and the trigger's role as a superset was highlighted; then both the trigger and referent would be marked as subset, and a trigger link made between each subset and the superset. If the trigger was not highlighted, and/or was said by the other speaker, it was not a subset.

5.6.2.4.4 adverbial The speaker used a focus-sensitive adverb, i.e. *only*, *even*, *always*, *especially*, *just*, *also* or *too* to highlight the word, and not another in a plausible set (which did not need to be explicit). The set could be abstract if the invocation of an alternative set was clear. In particular, annotators did not include instances of *just* used as a discourse filler or downplayer, e.g. "It's just so realistic". Annotators marked the focussed word as adverbial and the adverb the trigger. Again, if the adverb was highlighted, it was also adverbial. For example, in (5.14), B didn't even like the "previews" of 'The Hard Way', let alone the movie.

(5.14) (A) I like Michael J Fox, though I thought he was crummy in 'The Hard Way'.
(B) (I didn't *even* like) (the **PREVIEWS** on that)

5.6.2.4.5 answer This category was intended to capture narrow or wh-focus. However, we used a broad definition since wh-questions were uncommon in the corpus. A word was an answer if it, and no other, filled an open proposition set up in the context by either speaker; such that it would have made sense if they had only said that word or phrase. For example, in (5.15), A sets up the "bloom" she can't identify, and B answers "lily". The trigger was not marked.

(5.15) (A) Well everybody down here calls these flags... they get just one bloom... I'm not sure what they are called but ... they come in all different colours the blooms are on some of them is yellow, purple, white just all different colours
(B) (I'm going to BET you) (that is a **LILY**)

5.6.2.4.6 other The category other marked cases where the word was clearly kontrastive, but it did not fall into any of the types set out above. Annotators were told to use this sparingly when the markable was particularly salient. For example, in (5.16) the speaker clearly wishes to highlight that it was *Christmas Eve*, and not any other day, that they had forgotten, so it was not background. However, it was not contrastive, as there was no explicit trigger, nor was there an explicit superset of *all days*, so it was not a subset. There was no focus-sensitive adverb for adverbial, nor an open proposition set up for answer.

(5.16) (When I was a little KID) (I saw 'the INCREDIBLE JOURNEY') (on CHRISTMAS EVE) (and it was SO GOOD) (that I had FORGOTTEN) (it was **CHRISTMAS EVE**)

	Training		JA Finish		JK Finish	
	Blind	***Agreed***	***Blind***	***Agreed***	***Blind***	***Agreed***
All Categories	0.624	0.889	0.657	0.828	0.721	0.823
± kontrast	0.620	0.863	0.663	0.817	0.722	0.821
Total Markables	(1049)		(1177)		(1268)	

Table 5.3: Inter-annotator agreement figures for kontrast type and presence/absence (±) of kontrast using the kappa measure (κ). Agreement was measured between the two annotators after the completion of the training period (Trained) and when JA finished his annotations (JA Finish), as well as between JK and the author when JK finished (JK Finish). Figures are given for blind agreement (Blind), and following discussion of disagreements (Agreed).

5.6.2.4.7 background Finally, background was the opposite of the categories above. It was used for words that were either not salient, or salient with no implication of alternatives. This could be because the word related back to what had been said before. For example, in (5.16), "it was" in "it was so good" relates back to "the Incredible Journey", and "I had" to "when I was a little kid". The category was also used when the speaker introduced a completely new proposition, with no implication of contextual alternatives, e.g. in (5.17), "rats in the attic" is highlighted, but this is not differentiated from other things in the attic, or other places rats might be, etc.

(5.17) (I was living ALONE) (at the TIME) (and it was LATE at NIGHT and) (SCARY and) (you start HEARING NOISES) (and there's **RATS in the ATTIC**)

Annotators were told to expect that almost every sentence would contain at least one background, as well as some sentences which were entirely background. This rather broad definition of background probably led to some 'false positives', particularly the kontrast in 'all new', broad focus clauses was missed. It was difficult to see how to avoid this difficulty within our scheme, however.

5.6.3 Annotation and Annotator Agreement

The annotations were done by two paid post-graduate linguistics students at the University of Edinburgh, both with a general linguistics background. There was a fairly extensive training

Category	K	(T)
Correction	0.857	(21)
Contrastive	0.806	(851)
Subset	0.707	(638)
Adverbial	0.785	(99)
Answer	0.823	(17)
Other	0.647	(314)
Background	0.835	(4294)
Non-Applicable	0.941	(754)
Total	**0.845**	**(3494)**

Table 5.4: Reliability of the individual kontrast types in agreed annotations, kappa (κ) scores given, as well as the total number of times each type was chosen by either annotator (T) (Total is total number of markables).

and discussion period. JA was only available at the beginning, so completed 36 annotations, compared to 105 by JK, plus two by both for comparison. After the training period, annotations took an average of 4-5 hours per conversation.

At the end of the training period, annotators reported that they understood and felt confident about their task. Periodically, the annotators checked over each others' annotations, and recorded disagreements. Typically, these averaged 20 per conversation, reasonable given an average of 850 markables. Agreement was measured on three conversations (see Table 5.3): the first after the training period; the last two to check maintenance of consistency when JA and JK finished respectively (*JK finish* is compared to the author).[10] "Blind" agreement shows kappa without discussion. Given the level of confidence of the annotators, $\kappa = 0.62 - 0.72$ was lower than hoped. As noted above, Carletta (1996) reports this only merits "tentative conclusions". However, she also says that certain tasks, particularly discourse segmentation, may be inherently more difficult to annotate than those previously reported, such as word or clause boundaries. Being a new feature, it is difficult to know what a 'good' level of agreement is. We decided to analyse the disagreement further. For each conversation, annotators met with the author to discuss which disagreements were genuine, and in which, especially using the decision tree, they could agree on one or other annotation. This resulted in $\kappa = 0.86 - 0.89$. The areas of disagreement show the inherent difficulties in this

type of analysis.

For units larger than the NP, annotators marked the word or words that sounded salient (see section 5.6.2.1). This led to conflicts about both scope and salience. In (5.18), for example, the speaker is talking about how to deal with problem youth in the inner cities. Both annotators agreed that "I'm not good enough to raise my child" is a subset of "hostilities". However, one thought the most important element was *good enough*, while the other thought it was *raise my child*. Since the speaker stressed both *good* and *my child*, either or both annotations seem defensible.

(5.18) if they did have a big brother, big sister program... the parents might have *hostilities* towards them... like...
(I'm not *GOOD enough*) (to *raise MY CHILD*)
which basically is true

It is difficult to see how to avoid this type of ambiguity. Segmentation is the most difficult aspect of such annotation tasks (e.g. see the Prague annotation results).

Certain types were better identified than others (see Table 5.4). One reason for this was that there was often more than one plausible classification. In (5.19), for example, the speaker is talking about how hard it is for children of drug addicts. One annotator thought *leave* was a subset of *how*, i.e. ways to *make it better*, while the other didn't think this was plausible, and marked *leave* as adverbial because of *just*.

(5.19) I feel for them... I don't know *how* to make it better for them...
(they can't *just* LEAVE)
and say OK, well it's not acceptable

The decision tree (see Appendix D) helped in such cases. However, its efficacy depended on the annotators realising the competing analyses, and finding them both plausible. Many differences in the blind comparison were because analyses were not noticed; and many outstanding disagreements because the annotators did not find one or other analysis plausible. This accords with our understanding of ambiguity in language in general: on close analysis, much of language is highly ambiguous, something we normally do not notice as long as we can derive a sensible meaning.

The category of other was somewhat problematic. In particular, there were more other/background disagreements than with other types. This could be because there were two

[10]Note that agreement was checked for each *word*, e.g. if one annotator marked the NP *a good movie* as contrastive, and the other just marked *movie* as contrastive, that would be agreement on *movie* and disagreement on *good*.

Type	Word	NP	Ave. Lg	Freq
Correction	169	54	2.39	0.2%
Contrastive	6823	1885	2.30	7.8%
Subset	5037	2273	2.38	6.6%
Adverbial	1798	160	2.34	1.8%
Answer	196	116	2.43	0.3%
Other	6166	1544	2.32	6.9%
Background	91856	n/a	-	82.7%
Non-Applicable	13325	n/a	-	-
Total	**124440**	**6962**	**2.19**	**111115**

Table 5.5: Distribution of kontrast types at the word and NP level, and the average length in words at NP level over all 143 conversations annotated for kontrast (frequencies exclude non-applicables).

criteria: a clear alternative set and especial emphasis, and therefore more room for disagreement. For example, here the speaker is again talking about dealing with problem youth. One annotator thought *younger* clearly contrasted with the unstated *older*, where the other thought that it followed from a discussion about children and was not kontrastive.

(5.20) (it seems that the YOUNGER you can get them)
... involved with programs... you might keep them

This type of disagreement is expected given the nature of alternative set interpretation, see section 2.2.1.3. Further, as we saw in section 3.2.4, emphasis can be either categorical or gradient, and varies according to the speaker's level of involvement, e.g. bored or exaggerated. Overall, we decided it was better to keep the category, because of cases (such as (5.16) above) which were clearly kontrastive, but did not fit in one of the other types. The annotators' difficulty does vindicate our decision not to annotate kontrast per se, however.

In general, the annotations were reasonably successful, given the lack of precedent for annotating information structure in spontaneous English conversation. Further development of such a standard will want to look again at the issue of kontrast scope and the status of other.

5.6.4 Distribution of Kontrast Types

Table 5.5 shows overall frequencies of kontrast types. About 83% were annotated as background (excluding non-applicables). We saw in section 5.4 that around 60% of the words in the corpus are not accented. Therefore, this is consistent with our contention in Chapter 3 that only certain prosodic prominences mark kontrast. Unfortunately, there were not very many answers, showing again that classic wh-focus examples are uncommon in spontaneous conversation. Finally, we can see corrections, contrasts and others were less likely than subsets and answers to be marked at the NP level, while adverbial was usually at the word level (92%).

5.6.5 Kontrast Features

Once more, we used NXT queries to extract kontrast features from the annotations (see Appendix E). We distinguished kontrasts marked at the word-level and NP-level from backgrounds; as well as kontrast boundaries. There were also features trying to capture the effect of kontrasts on each other in context, including the distance to the trigger, or to other kontrasts in the clause or phrase. Lastly, we noted above that annotators were told not to mark kontrast in all-new sentences with broad focus. This meant that kontrast could not be predicted in such clauses. We therefore decided to exclude words in clauses which did not contain at least one kontrast (26.7%).

5.7 Theme and Rheme Annotation

In Chapter 4 we concluded that kontrast within themes is distinguished from kontrast within rhemes by prosodic subordination, not a specific pitch contour (e.g. L+H* LH% versus H* LL%) as claimed by Steedman (2000) and others. Here we wanted to test this claim on our corpus. However, even more so than with kontrast, it is hard to determine the scope of the theme/rheme units. Themes are often not prosodically prominent and are indistinguishable (both theoretically and practically) from non-kontrastive elements in rhemes. As we saw in section 5.6.1, there is little prior work to guide us on this. But while there is no clear syntactic basis on which to determine scope of the theme/rheme units, there is more widespread agreement that they can be marked by phrase boundaries (see discussion in Chapter 3 and Kruijff-Korbayová & Steedman 2003). Therefore, we annotated prosodic phrases for theme/rheme status. We defined this in terms of a positive test for themehood. Any phrase that was not a *theme* would be marked as a *rheme*, so rheme phrases might contain backgrounded thematic material. A *theme* phrase only contained information which linked the utterance to the pre-

ceding context, i.e. setting up what the speaker was saying in relation to what had been said before. For example, here *my area* is already established, and so it is thematic. (Kontrasts are also shown, however kontrast annotation was entirely separate).

(5.21) (Q) Personally, I love hyacinths.
What kind of bulbs grow well in your area?

(A)	(In	MY	AREA)	
	Bkgd	*Kont.*	*Bkgd*	*(Theme)*
	(it is the	DAFFODIL)		
	Bkgd	*Kont.*		*(Rheme)*

The complication with this was that the prosody annotation had to be completed first. Therefore, within the scope of time and funding for this thesis, only one conversation has been annotated, by the author, for theme/rheme status.[11] All phrases with 2p, 3 and 4 breaks were included to maximise the amount of data. However, phrases which the author felt were too disfluent for the information structure to be clear, or which contained only non-propositional content, e.g. *anyway*, were excluded. In total there were 110 theme phrases and 184 rheme phrases. The author also indexed theme and rheme phrases which were clearly part of the same information unit (as in the example above). There were 50 such paired information units. As before, these annotations were used to extract features, including *theme/rheme status*, *place* and *order* (see Appendix E).

All of the features described above are included in the data set used for building models to test our predictions about the relationship between prosody and information structure in the next two chapters. The experiment data set only includes words for which all of these features could be extracted, i.e. minus the exclusions noted. Such exclusions are an inevitable part of trying to test high-level semantic theories on a unconstrained corpus of spontaneous speech. Overall, with the integration of all the existing annotations of the Switchboard corpus, plus the addition of substantial new annotations for prosodic information and kontrast; this corpus is one of the richest resources available for the study of information structure in English.

[11] The prosody and kontrast annotation for this conversation were done by others to make it as unbiased as possible.

Chapter 6

Predicting Prosodic and Information Structure

In Chapter 3 we argued that prosodic prominence is defined on metrical structures, rather than resulting from accenting per se. That is, the perception of the prominence of a word is mediated by the phrasal structure in which it appears, rather than resulting solely from the acoustic properties of a word itself. In particular, we presented previous work that showed that the *nuclear* accent is not necessarily the most acoustically prominent accent in a phrase; but is rather the head of the strongest node in the metrically branching structure of the phrase, which by default is right-branching. In section 3.2.1, we claimed that one of the functions of this prosodic structure is to align the heads of information structural units, i.e. kontrastive elements, with positions of nuclear prominence. However, this would only occur if the information unit was "heavy" enough to form its own prosodic phrase. "Weight" is determined by a combination of semantic and phonetic factors. These include information status, i.e. rhematic elements are more likely to form a phrase than thematic ones; discourse status, i.e. given elements are "lighter" than new; and syntactic structure, i.e. certain types of clause and constituent are more likely to be contained in their own phrase; as well as phonetic factors, including phrase length, speech rate, rhythm and emphasis. In addition to this, the acoustic prominence of either a phrase or an accent can be independently manipulated to emphasise that phrase or word. Acoustic prominence interacts with structural prominence leading to the interpretation of kontrast status and information structure. Section 3.3.1 discussed how these interacting factors could be modelled. We argued that metrical prosodic structure is both produced and parsed probabilistically in terms of the most likely alignment of a branching prosodic structure with the segmental string. The intended prominence of the nodes in that structure is interpreted given the acoustic, semantic and phrasal properties of the words in the string.

In Chapter 4, we saw how this alignment works to convey theme/rheme status at the inter-phrase level. In the experiments in this chapter, we concentrate on the intra-phrase level. We show how the claims above are consistent with the properties of our corpus, by investigating the semantic, syntactic, phrasal, accentual and acoustic features that are useful to predict the basic elements of these interacting structures, i.e. phrases, accents, kontrast and prominence. Finally, we look once more at the properties of theme/rheme phrases, to consolidate the result in Chapter 4. For each element, we will begin by setting out the claims of our theory in relation to that element, before showing how the results from our prediction models for that element are consistent with these claims. Firstly, we will briefly review related work on the prediction of accents and phrase breaks. We then describe the statistical classifiers used to build the prediction models. In the models described in this chapter, quite a lot of importance is put on the relative effectiveness of different *types* of features, therefore, we further begin by set out the justification for these feature set groupings.

6.0.1 Related Work

The studies most closely related to those reported in this chapter are those which have tried to predict prosodic events (usually pitch accents and phrase boundaries) automatically from hand-labelled speech corpora. The ultimate aim of most of these projects has been to increase event recognition accuracy in order to improve automatic speech recognition, natural language understanding or the input to speech synthesis systems. While it is hoped that the current work will contribute to this aim, its major purpose is to assess our theoretical work on the importance of different types of features used for predicting prosodic events; and in particular, what prosodic events are most important to convey meaning. Therefore, we have not tried to optimise accuracy scores for the current prosodic event prediction tasks, in terms of classifiers used, classifier-internal parameters, and acoustic and syntactic feature tuning. Here we briefly present representative previous work, in order to show that the performance of our own prosodic event prediction models is comparable, and therefore that our models do a reasonable job of explaining the data; without claiming that our models, in their current form, will lead to improvements in these natural language systems.

Hirschberg (1993) reports pitch accent prediction accuracy on a number of different corpora. The most similar to ours is a collection of multiple-speaker 'wizard-of-oz' air travel requests (i.e. a human talking with a computer controlled by a human). She reports accuracy of 81.9% using part-of-speech and limited discourse status (given/new) features, rising to 85.1% when phrase position features were included. Conkie et al. (1999) report accuracy of 84.0% using part-of-speech features, 82.8% using localised pitch and intensity features, rising to 88.3% combined on a corpus of single-speaker read newspaper text. Taylor's

(2000) prosodic event detection system detected 71.5% of all prosodic events (accents and boundaries) by syllable in the Switchboard subset described in section 5.4.1, using only local pitch and intensity features. He does not report accuracy for accents and boundaries separately. Pan et al. (2002) investigate the effect of different types of features on prosodic event prediction in a corpus of medical spontaneous monologues and read speech. They report accuracy of 84.6% for their final model, showing that part-of-speech, syntactic, discourse status and semantic features such as informativeness all significantly improved performance. Finally, Chen & Hasegawa-Johnson (2004) report accuracy of 84.2% on a limited-speaker radio news corpus, showing that local acoustic features led to little improvement over their part-of-speech/syntactic features.

To our knowledge, there has been less work on phrase break prediction. The latter two studies also report accuracy for phrase boundary detection. Pan et al. (2002) report recognition of level 3 breaks as 89.4% and level 4 breaks as 90.4% using the features described above; showing that syntactic constituent features were more important than the semantic informativeness and discourse status features used for pitch accent prediction. Chen & Hasegawa-Johnson's (2004) model achieved accuracy of 93.1% on major intonation phrase boundaries (level 4 break), again showing that semantic features were much more important than local acoustic ones. Shriberg, Stolcke, Hakkani-Tür & Tür (2000) report results from a 'sentence' boundary detection task on a one million word subsection of Switchboard. They used automatically extracted acoustic features against a statistical language model. These are the same units used for dialog act classification (described in section C.2) and are often considerably longer than prosodic phrases. They report an error rate of 4.0 compared to a baseline of 11.0, with the acoustic features used showing little improvement over the language model.

As can be seen, reported accuracy on prosodic event classification is reasonably good, with the pitch accent detection task proving to be harder than phrase break detection. However, most of these studies use corpora from single or limited speakers, and the speech is either read, or limited domain and highly constrained. The findings of the two studies using the Switchboard corpus are hard to compare with our work, because their tasks are considerably different to ours. Most of the studies use very localised features, i.e. part-of-speech and word-level acoustic features. It may be that these are only as effective as they seem to be in such highly restricted genres. It is also unclear whether pitch accents, as defined in these studies, are actually adequate to convey meaning prosodically, as is the ultimate aim of these systems.

6.0.2 Classifiers

In the first two groups of experiments below, we will be using two statistical classifiers to build the models we are interested in. This was done in order to be able to assess more clearly how robust the effects we were finding were, and how much due to the properties of thc classifier. These two classifiers were used firstly because they model data quite differently: logistic regression models features as weighted effects over all data points; whereas CART, by its nature, picks out effects on clusters of features. Secondly, both build models that are easily interpretable in terms of the usefulness of different features. This was important, since it was the impact of different features and feature sets, rather than overall performance levels, that we were most interested in. Here we briefly set out how the different classifiers work, and how their output will be used to interpret results in our experiments.

6.0.2.1 Logistic and Linear Regression

In most of the experiments below, we will be using logistic regression models to predict the probability of different outcomes of various categorical variables. Logistic regression uses the same principles as linear regression, except that the former deal with categorical dependent variables, and the latter with continuous ones. Linear regression is based on the idea that the value of a dependent variable (DV) can be predicted from the sum of weighted factors (f) that affect that variable, plus a constant showing the initial value of the variable, and a term to describe the error, i.e.:

$$DV = \beta_0 + \beta_1 f_1 + \beta_2 f_2 + \beta_3 f_3 + ... + \beta_n f_n + \epsilon_i \tag{6.1}$$

In logistic regression, instead of predicting the value of the dependent variable, the *probability* of a particular outcome (Y) is predicted using the following equation:

$$\begin{aligned} P(Y) &= \frac{1}{1+e^{-Z}} \\ Z &= \beta_0 + \beta_1 f_1 + \beta_2 f_2 + \beta_3 f_3 + ... + \beta_n f_n + \epsilon_i \end{aligned} \tag{6.2}$$

Broadly, this applies maximum likelihood estimation (MLE) after transforming the dependent variable into a logit variable. This is the natural log of the odds of the dependent occurring or not. MLE tries to maximise the *likelihood ratio* (-2LL) of the model, i.e. the odds that observed values of the dependent may be predicted from the observed values of the independent variables. This likelihood ratio is therefore a measure of how well the model fits the data overall. Since the distribution of likelihood ratio values is roughly chi-squared, a chi-squared test can be used to assess whether the addition of any variable significantly

improves the amount of variation in the dependent that the model explains. This test can be utilised to exclude non-significant variables from the model. In many of the models below, experimenter-controlled backward stepwise logistic regression is used to remove redundant variables. That is, all variables are entered in the model, and then each is excluded at a time to see whether its exclusion significantly harms the performance of the model (in terms of the likelihood ratio); while the experimenter carefully controls the relative effects of similar features.

Instead of β-coefficients, the weights on factors in the model are logit coefficients (B). These are the natural log of the odds ratio for that variable. The most common way to interpret these is to convert them back to the odds ratio (*Exp(B)*). This ratio shows the effect on the odds of the outcome for that variable. Since the odds is the probability of a given outcome occurring over the probability of it not occurring, a value of 1 shows the variable has no effect on the outcome. Values between 0 and 1 show a negative effect, and values greater than 1 a positive effect. For categorical variables, this is the effect on the odds for a variable having a particular level (compared to the mean over all levels). For continuous variables, it is the effect of a one unit increase in the variable. To ease interpretability, Exp(B) can be used to calculate the percentage difference (*Pdiff*) in the probability of an event with the inclusion of each variable, controlling for all the other variables in the model. This is done by calculating the new odds of an outcome ($Exp(B_y)$) by multiplying the original odds ($Exp(B_0)$) by the odds ratio for the variable ($Exp(B_x)$). The new odds are then used to find the probability of the outcome with that variable (P_y). The difference is this value minus the prior probability of the outcome (P_0), i.e.:

$$\begin{aligned} Exp(B_y) &= Exp(B_0) * Exp(B_x) \\ P_y &= \frac{Exp(B_y)}{1 + Exp(B_y)} \\ P_{diff} &= P_y - P_0 \end{aligned} \tag{6.3}$$

Most of the models below involve binary logistic regression, i.e. the dependent variable is dichotomous. The model predicts the likelihood of the outcome of interest in relation to the other, reference, value. In A3, we report one set of models using multinomial logistic regression, i.e. the dependent has more than two levels. In this case, each level is compared to the reference value. In the final experiments, we are interested in the effect of different variables on several continuous dependents, and so use MANCOVA testing. This employs a widely-used form of linear regression model where dummy variables are used to model the effect of categorical independent variables on multiple continuous dependent variables.

```
((propSyl_ph < 1)
 ((kon_stat is backgd)
  (((noaccq 0.915523) (nuc 0.0844765) noaccq ))
  ((kon_stat is konnp)
   ((accsPh_exc < 0.6)
    ((dur_relSyl < 0.152023)
     (((noaccq 0.69697) (nuc 0.30303) noaccq))
     (((noaccq 0.255319) (nuc 0.744681) nuc)))
    (((noaccq 0.756579) (nuc 0.243421) noaccq)))
   ((dur_relSyl < 0.189221)
    ((propSyl_ph < 0.6)
     (((noaccq 0.694444) (nuc 0.305556) noaccq))
     (((noaccq 0.405797) (nuc 0.594203) nuc)))
    (((noaccq 0.228916) (nuc 0.771084) nuc)))))
 ((kon_stat is konword)
  (((noaccq 0.0576923) (nuc 0.942308) nuc))
  ((accsPh_exc < 0.8)
   (((noaccq 0.0338983) (nuc 0.966102) nuc))
   ((POS_gp is NN)
    (((noaccq 0.269076) (nuc 0.730924) nuc))
    ((POS_gp is PR)
     (((noaccq 0.864407) (nuc 0.135593) noaccq))
     ((accsPh_exc < 1.6)
      ((POS_gp is VB)
       (((noaccq 0.5) (nuc 0.5) noaccq))
       (((noaccq 0.333333) (nuc 0.666667) nuc)))
      (((noaccq 0.705882) (nuc 0.294118) noaccq)))))))))
```

Figure 6.1: Example CART tree, used to classify nuclear accented words from unaccented words, including all features (described in section 6.2.1, see Appendix E for description of features).

6.0.2.2 CART

The other classifier used is the classification and regression tree (CART) (Breiman, Friedman, Olshen & Stone 1984). The implementation used in this project is the *wagon* CART

building program that is part of the University of Edinburgh's Speech Tools Library (King, Black, Taylor, Caley & Clark 2003). The tree is a binary decision tree used to assign a class to each data point. The tree asks a series of yes/no questions about the features of each data point in order to classify it. In the case of categorical variables, this is whether or not the datum has a particular level of that variable. Continuous variables are binned in order to be treated in the same way. In all experiments below, 5 bins were used. Figure 6.1 shows an example tree, used to classify nuclear accents from unaccented words including all features (see Table 6.6). As can be seen, these trees are readable, and can be used to assess the impact of different features.

In our study, the trees are derived automatically from the training data. At the beginning all the data is put at the root of the tree. The program then asks all possible questions about the features of the data set, selecting the one that splits the data so that each new set has the least impurity. This continues until all the data points at one node are the same class, or there are a minimum number of data points at that node, whichever comes first. In all experiments below, the minimum was 25. Impurity was measured in terms of the entropy of each set multiplied by the number of data points. Entropy is calculated by:

$$H = \sum_{x} P(x)log(P(x)) \tag{6.4}$$

where $P(x)$ is the probability of a data point x having a certain label, given its features, summed over all xs in the set. The number of bins for continuous variables, and the stopping level at each node can be tuned to maximise results for each task and data set. (Although, as discussed above, this was not carried out here).

For the models built using CART, results were tested using five-fold cross validation, as this method has been found to lead to more reliable results than a simple training/testing division when the sample size is small (Bailey & Elkan 1993). The data set was divided into five blocks with equal numbers of words in each. Each model type was then derived five times, each time with a different block as a testing set and the other four used to train the model. The results from the five runs were then averaged to get the final results.

6.0.3 Feature Type Groupings

In the discussion of the various prosodic and kontrast element prediction models below, there is considerable emphasis on the relative importance of different feature groups. Features are grouped according to the type of information they provide. The first group was *syntax/semantic* features. These were the inherent features of words which we have argued act as constraints on the alignment between words and prosodic structure. There were 11 dis-

course semantic features, including kontrast and information status, and kontrast boundary; grouped with 11 syntactic features, such as clause and constituent type and boundary. The next group is features of the prosodic structure (*phrasal* features), including the relationship between the current word and the structure. These were meant to capture the constraints on prosodic structure, as well as acoustic properties of each phrase overall. There were 14 phrasal features. In some of the models these were divided into *positional* features, which encode the position of the word in relation to the phrasal structure, e.g. duration of the phrase so far, and number of syllables in the phrase so far; and *whole phrase* features, such as speech rate, normalised mean pitch and 'eurthymic' features, i.e. comparing the phrase to the previous phrase, e.g. number of syllables in the previous phrase. The former were meant to encode the phrasal constraints on the current word, while the latter were meant as control variables on the occurrence of accents overall, e.g. when speaking more slowly speakers tend to accent more words. In the first set of models *accentual* features, e.g. whether the word is an accent, and the number of accents in the phrase so far, are further separated from the rest of the phrasal features; as we were trying to assess how much accenting predicts phrasing. In the rest of the models, these are included with the positional features as they are manifestations of the preceding prosodic structure. The final group of features was *word-level acoustic*, such as the normalised pitch, intensity and duration of the word itself. We viewed these features as *manifestations* of prosodic structure, rather than constraints upon it, e.g. a word in a prominent position in the structure is likely to be longer, louder, etc. Therefore, these features were added to the models last, as a way of assessing how well the other features could model each element. That is, the smaller the improvement in recognition, the more useful the other features, because the word-level acoustic features did not add very much more information. In total there were 47 possible features (see Appendix E.2.1 for full list).

6.1 Phrase Breaks

Our theory of the relationship between prosody and information structure gives central importance to the role of the nuclear accent. However, we define the nuclear accent as the strongest node in the metrical tree of its phrase, which is by default right-branching. The consequence of this, in terms of predicting the basic elements of prosodic structure, is that the prediction of phrase breaks, i.e. the division of the speech signal into prosodic phrases, can be said in some sense to come "first". By this, we mean that phrasal structure can determine the perception of prominence, especially nuclear prominence, rather than the other way around. We do not mean to make any specific claims in terms of a generative model

of language production. Since we also claim a strong constraint aligning nuclear accents and kontrast, it follows that prosodic phrasing is a strong constraint on information structure. Since the units of information structure are broadly syntactic, i.e. elements of a proposition, we expect syntax to constraint phrasing; although, as we have seen in many examples to this point, information units do not necessarily align with traditional syntactic constituents. Of course these constraints are probabilistic, so we expect them to interact with general phonetic constraints on phrasing such as speech rate and emphasis.

In this experiment, we wish to show that this general model works on real speech data in terms of the most basic claim it makes, i.e. that all of the factors discussed above affect phrasal structure. We test this using a model which predicts whether a word is followed by a phrase break. Our aims are two-fold: firstly to assess the importance of the different factors discussed in Chapter 3, and show that they are consistent with our general model; secondly, to show that accents have only a small effect on the likelihood of a phrase break, consistent with our "structure first" view.

6.1.1 Aim and Method

Our general claim and the specific hypotheses being tested in this experiment are therefore:

- **General Claim:** Prosodic phrasing results from probabilistic mapping between metrical prosodic structure, syntactic structure and information structure. The head of an information structure unit "wants" to map to the head of a phrase, but can only do so if the syntactic and phonetic properties of the unit are sufficiently "heavy". Since prosodic prominence can therefore largely be defined structurally, phrasing can be thought of as coming "before" accenting.

- **Hypothesis 1:** Phrase breaks are most effectively predicted by a combination of kontrast and prosodic and syntactic features.

- **Hypothesis 2:** Features related to the distribution of accents in the phrase (accentual features) do not substantially improve the accuracy of phrase break prediction.

Our method was to build models which predict the probability of each word being followed by a phrase break (break index 3 or 4), using the features described in the last chapter. These models were built using different combinations of feature types as set out below. Two different classifiers were used, CART and logistic regression, as described in section 6.0.2. The efficacy of the different features was determined by how accurately each feature could predict phrase breaks, and the overall importance of each feature in the different models.

The data set consisted of the words from the 18 Switchboard conversations described in section 5.4. These conversations were all annotated with prosodic features according to our annotation scheme. Only words which had been annotated for kontrast status were included. As well, words which were in syntactic clauses which did not have at least one kontrast were excluded, as discussed in section 5.6.5. Words which occurred in disfluent phrases, i.e. with a *2p* or *X* boundary were also excluded as set out in section 5.4.2.1. The semantic, syntactic, prosodic and acoustic features described in the last chapter were extracted for these words (see Appendix E.1 for a description of all the extracted features). A small percentage of words then had to be excluded because one or more of its syntactic features were not found (as discussed in section 5.3); or because the pitch and intensity either could not be extracted by Praat, or were excluded by our normalisation procedures (as discussed in section 5.5). After this processing, the final data set was 8915 words from 33 speakers.

A phrase break prediction model was then built for each feature type using each classifier. For the CART classifier, all features for each group were added at the beginning, and then removed one by one if they either did not improve, or harmed, the number of breaks correctly classified. For the regression classifier, all features were added, and then the backward stepwise logistic regression was used to exclude non-significant features (see description above). In both cases, the process was controlled by the experimenter to specifically assess the contribution of similar features, e.g. position in the phrase in syllables versus position in the phrase in words. For each classifier, features which were significant in the type model were then included in the combined feature type models and the feature reduction process was repeated, so that all reported models only include features which significantly improve the accuracy of that model (see Appendix E.2.1 for full list).

6.1.2 Results and Discussion

Table 6.1 shows the performance of each classifier for each group of features, in terms of the percentage of words correctly classified as being in a phrase, at a phrase break, and the overall percentage correctly classified. This is compared to baseline performance, where each word is simply classified as being the most likely category, i.e. in a phrase. Recall that the CART classifier performance is based on five-fold cross validation, while the regression model is trained and tested on all the data. The accuracy of the regression models may therefore be slightly inflated, although the effect seems to be relatively minor. We also report the likelihood ratio and chi-squared significance tests comparing the amount of unexplained variance in each model to the last for the regression models (see section 6.0.2). This gives a more reliable indicator of performance than classification tables since it accounts for the actual probability measured for each outcome, rather than just whether the probability is

PhBrk	CART Classifier			Regr Classifier			Regr Model Fit		
	InPh	PhBk	Acc	InPh	PhBk	Acc	-2LL	Sig	χ^2 (df)
Baseline	100	0	70.6	100	0	70.6	12275	-	-
Phr	87.4	59.0	79.2	90.7	55.7	80.4	8413	.000	3861 (13)
Sem	92.2	60.6	83.1	92.5	64.4	84.2	3993	.000	4402 (40)
Sem+Phr	92.3	66.2	84.8	93.4	72.3	87.2	3264	.000	1736 (16)
Sem+Phr+Wd	-	-	-	93.5	73.4	87.7	2825	.000	373 (3)
Sem+Phr-Acc	91.6	65.6	84.1	93.3	70.8	86.7	3464	.000	287 (2)

Table 6.1: Phrase break prediction for all 8915 words by classifier (CART versus logistic regression (Regr)). Percentages of words correctly classified as being in a phrase (InPh), before a phrase break (PhBk) and overall accuracy (Acc) are shown for models built using different combinations of features: phrasal (Phr); semantic (Sem); combination semantic and phrasal (Sem+Phr); semantic, phrasal combined with word-level acoustic (Sem+Phr+Wd); and the semantic/phrasal model excluding accentual features (Sem+Phr-Acc). The likelihood ratio (-2LL) and chi-squared significance tests are also given for the regression models. See text for more details.

above or below 0.5.

In relation to the first hypothesis, we can see that these results are consistent with our claims. Each group of features (semantic/syntactic (Sem) and phrasal (Phr)) improves performance over the baseline; and, for the regression model, significantly decreases the unexplained variance. Furthermore, adding the syntactic and phrasal features together improves performance, and the chi-squared test is significant (χ^2 figures for Sem+Phr compare the Sem and Sem+Phr models). Adding word acoustic features also slightly improves performance (in the regression model). This model comparison provides basic support for our hypothesis. In order to take the point further, we will look more closely at which features were significant in the final model, and their contribution, below. In relation to the second hypothesis, we can see that in the regression model, the inclusion of accentual features does significantly improve the model (χ^2 test shows the effect of adding accentual features to a Sem+Phr model without them). However, the improvement in accuracy for both classifiers is small. For the CART model, exclusion leads to the percentage of breaks correctly classified going down 0.6%, 2.6% for the regression model. In both cases overall accuracy goes down less than 1%. Therefore, overall, accentual features do improve the accuracy of phrase break

Feat	Exp(B)	P diff	Sig	Wald (df)
repair	3.77	31.7%	.001	11.1 (1)
numWd_cl	1.03	0.7%	.000	40.2 (1)
clbound	4.76	37.1%	.000	241.1 (1)
cnsbound	4.04	33.3%	.000	43.2 (1)
ident_inf	1.98	15.8%	.000	21.1 (1)
bound_inf	2.94	25.6%	.008	6.9 (1)
event_inf	1.65	11.4%	.003	8.9 (1)
NN	1.45	8.3%	.019	5.5 (1)
cnsbound by adjunct	2.36	20.2%	.000	31.5 (1)
cnsbound by obj	1.56	10.0%	.002	9.7 (1)
t_ph	1.62	10.9%	.000	299.6 (1)
accsPh_inc	1.91	14.9%	.000	76.4 (1)
anyaccq	1.70	12.0%	.000	31.3 (1)
posWd_ph by spRate_syl	1.15	3.0%	.000	202.5 (1)
adjunct by posWd_ph by propWd_cns	1.05	1.1%	.014	6.0 (1)
dur_relSyl	1.97	15.6%	.000	285.9 (1)
npqrange_wd	1.10	2.0%	.000	16.5 (1)

Table 6.2: Factors which significantly increase the likelihood of a phrase break in the full regression model (i.e. Sem+Phr+Wd, see Appendix E for description of features). The odds ratio for each feature in the model (Exp(B)), as well its significance (Sig) using the Wald statistic is given, along with the percentage increase (P diff) in the likelihood of a phrase break with the presence of that level of that variable (or a one unit increase for continuous variables).

prediction, but not substantially. We will return to this point below.

Overall, accuracy of the final model was 87.7% for the regression classifier and 84.8% for CART. This is lower than the figures reported in section 6.0.1, which ranged from 89.4-93.1%. However, as we said there, those results were on monologue and read speech respectively. Such speech tends to have longer phrases that are more carefully controlled, and therefore breaks probably align more consistently with syntactic phrase boundaries. Although disfluent phrases were excluded here, spontaneous speech still includes more planning pauses mid-sentence, and a more irregular speech rate.

Feat	Exp(B)	P diff	Sig	Wald (df)
inkon	0.68	-7.3%	.018	5.6 (1)
head_cns	0.74	-5.9%	.000	12.3 (1)
rel_inf	0.10	-25.6%	.000	17.4 (1)
nextold	0.69	-7.1%	.000	22.7 (1)
nextmed	0.83	-3.8%	.015	5.9 (1)
propWd_cns	0.20	-21.7%	.000	26.7 (1)
cnsbound by subj	0.32	-17.5%	.000	25.3 (1)
sylRel_lastPh	0.93	-1.5%	.006	7.6 (1)
posPho_ph	0.85	-3.2%	.000	64.4 (1)
posWd_ph	0.31	-18.0%	.000	223.7 (1)
posWd_ph by wd_lastPh	0.99	-0.3%	.001	10.2 (1)
subj by posWd_ph by propWd_cns	0.89	-2.3%	.000	14.3 (1)
npquan_wd	0.91	-1.9%	.000	44.3 (1)
Constant	0.06	-	.000	61.0 (1)

Table 6.3: Factors which significantly decrease the likelihood of a phrase break in the full regression model (i.e. Sem+Phr+Wd, see Appendix E for description of features). The odds ratio for each feature in the model (Exp(B)), as well its significance (Sig) using the Wald statistic is given, along with the percentage decrease (P diff) in the likelihood of a phrase break with the presence of that level of that variable (or a one unit increase for continuous variables).

Tables 6.2 and 6.3 show the variables which significantly increase and decrease the probability of a word ending in a phrase break respectively. Only significant variables are shown, and in the case of categorical variables, only the significant levels of that variable (see Appendix F.1 for the full listing). For each variable, the odds ratio (Exp(B)) is reported (see description in section 6.0.2), along with the significance of the variable using the Wald statistic. We also report the percentage difference in the probability of a break (see section 6.0.2). For durational features, one unit is 10ms; for pitch features, one unit is a 10% increase in the speaker's logged pitch range.

These results immediately show us the importance of syntactic phrasing on prosodic phrasing. The feature values clbound and cnsbound both greatly increase the probability of a phrase break. In fact, overall 72.3% of clause boundaries coincide with phrase boundaries. We can see that this is mediated by constituent type and structure: adjuncts (adjunct), objects (obj) and nouns (NN) are more likely to be followed by breaks, subjects (subj) and

constituent heads (head_cns) less likely. These constraints concur with generally understood properties of syntax structure. There is no absolute correspondence, but the relationship can be sucessfully modelled in terms of probabilistic constraints on phrase breaks. We can see a small effect of kontrast status: words that occur in a kontrastive phrase (inkon) are less likely to be followed by a break, however the decrease in probability is only 7.3%. This is probably because kontrast status only indirectly marks information units, as many kontrasts are single words which occur in the middle of a phrase over which that kontrast has scope. As discussed in section 5.7, we could not capture information unit boundaries directly, and therefore their constraint on phrasing. Finally, we see small effects of information status, e.g. if the next word is old (nextold) or mediated (nextmed) a break is less likely, showing the upcoming element does not have enough weight to form a phrase.

Constraints on phrasing structure itself can also be seen. As would be expected, a phrase break is more likely the longer the duration of the phrase so far (*t_ph*). The likelihood of a break increases with position as the speech rate increases (*posWd_ph* by *spRate_syl*). It also increases with position as the phrase gets towards the end of an adjunct (adjunct by *posWd_ph* by *propWd_cns*). The interaction of these features shows again that there is no absolute constraint of phrase length, or syntactic constituent, on phrasal structure; more the interplay of different probabilistic constraints determines prosodic phrasing. We can see some indication that rhythm has a measurable effect on phrase break position. For an utterance to be rhythmical we expect strong beats to occur at approximately equal intervals, and therefore that there will be an effort to make consecutive phrases have roughly the same number of beats. Here we see that the position of the current syllable relative to the number of syllables in the previous phrase (*sylRel_lastPh*), and the position of the current word by the number of words in the previous phrase (*posWd_ph* by *wd_lastPh*) both decrease the likelihood of a phrase break. The effect is small, but is consistent with a constraint against the current phrase being metrically longer than the previous one.

We can see that both acoustic word features, and accentual features, work in the direction that would be expected, e.g. the duration of the word (*dur_relSyl*) and the number of accents in the phrase so far (*accsPh_inc*) increase the likelihood of a break. However, as we saw in Table 6.1, the effect of both these set of features on overall performance is small. It seems reasonable to take this as an indication of the success of the model, as well as confirmation of our second hypothesis. We are claiming that features such as final lengthening and accenting are manifestations of phrase structure, rather than constraints upon it. That these features had little independent effect indicates that the other features in the model do indeed predict phrase boundaries successfully, making the word-level acoustic and accentual features redundant. We will see this more clearly in our nuclear accent prediction models in later sections.

In summary, our classification results showed that phrase breaks are indeed most effectively predicted by a combination of syntactic, phrase level and semantic features. Our analysis of the significant features in the final model showed that syntax, i.e. clause and constituent type and structure, acts as a strong constraint on phrasing. However, this is mediated by positional and rhythmical constraints on the prosodic structure itself; and, to a less demonstrable degree, by kontrast and information status. These results can be seen as consistent with the view of prosodic phrasing set up in Chapter 3, and as providing the foundation on which our claims about the perception and role of prominence, particularly nuclear prominence, lie.

6.2 Accents

In section 3.1.2, we laid out evidence from the literature that accents are a manifestation of strong nodes in a binary-branching metrical prosodic structure, rather than being independent events that can fall on any word in a phrase. A nuclear accent is perceived as the most prominent in a phrase because it is in the most structurally strong position, which by default is toward the end of the phrase. Other 'accents' appear as required on syntactically 'strong' elements to preserve the rhythm of the phrase, i.e. approximate isochrony between strong beats. This structure is manipulated to convey information structure by a constraint aligning kontrast with nuclear prominence. In terms of an accent prediction model, therefore, we expect distinct distributions of the properties of nuclear and plain accents. Nuclear accents should be defined in terms of prosodic structure, and not necessarily their acoustic properties. Plain accents, on the other hand, may appear for reasons less connected to information structure, and so may only be identifiable in terms of their acoustic properties. Nuclear accents are more likely to be directly 'meaningful', i.e. kontrastive. Plain accents will be less able to be predicted by information structure, but may attract prominence because of inherent properties such as part-of-speech type. As we discussed in section 3.2.2, these constraints are probabilistic: some plain accents may be 'meaningful' in this sense, e.g. a short, given kontrast may not form its own phrase; on the other hand, a nuclear accent may not be kontrastive if the constituent it occurs on is long enough to be in its own phrase.

In the following experiments, we test these claims on our corpus. In the first experiment, we look only at accented words. We wish to show that, as claimed above, nuclear accents can be distinguished from plain accents by their structural properties, both in terms of their place in prosodic structure, and in syntax/semantic structure. In the second experiment, we test the claim that nuclear accents are more directly 'meaningful' than plain accents. We do this by examining the factors that can reliably distinguish words with plain accents from

words with no accent to the factors that distinguish nuclear accented from unaccented words. Finally, we present an overall accent prediction model. This will be used to show once more the interacting constraints on prominence prediction; and to confirm the hypothesis in the last experiment that phrasing 'comes first'.

6.2.1 A1: Nuclear Accents are Structural

The first experiment therefore looks at the features which distinguish nuclear accents from plain accents.

6.2.1.1 Aim and Method

Our general claim and the specific hypothesis being tested in this experiment are:

- **General Claim:** Nuclear accents are primarily defined, and perceived, by their place in metrical prosodic structure. Specifically, a nuclear accent is usually the right-most strong node in a phrase. Therefore, we expect nuclear accents to fall towards the end of prosodic phrases; and, because of the mapping with syntax and information structure, toward the end of syntactic and information units.
- **Hypothesis:** Nuclear accents are most effectively distinguished from other accents by a combination of kontrast and syntax, and prosodic phrase position features.

Our method was to build models to predict the probability of a word carrying a nuclear accent, as opposed to a plain accent. We therefore excluded unaccented words from the data set. Nuclear accents were taken to include PN accents, and plain accents to include Q accents, as discussed in section 5.4.2.2. Again both CART and logistic regression classifiers were used, with different combinations of feature sets.

The data set was the same as that for Experiment 1, but with unaccented words excluded. There were a total of 4810 words. Features used were similar to the first experiment, except that full phrasal features were included, e.g. the number of words in the phrase and the number of words in the phrase so far relative to the number in the whole phrase (the proportion). In total there were 16 semantic/syntactic features. Phrasal features were divided into those that related to the position of the word in the phrase, e.g. accents so far or the proportion of syllables in the phrase so far (12 features); and those that did not, e.g. normalised mean pitch of the phrase or the number of words since the last accent (6 features); as we wanted to independently assess the usefulness of each type. The same 8 word-level acoustic features were also included, making a total of 42 features. Finally the same feature reduction process was carried out as in the first experiment, so that for each model for each classifier only features

N v A	CART Classifier			Regr Classifier			Regr Model Fit		
	A	N	Acc	A	N	Acc	-2LL	Sig	χ^2 (df)
Baseline	0	100	58.4	0	100	58.4	7274	-	-
Phr incl Pos	80.0	81.5	80.9	78.8	83.4	81.5	4448	.000	2826 (8)
Sem	61.7	70.7	66.9	56.0	78.6	69.2	6236	.000	956 (15)
Sem+otherPh+Wd	63.4	73.6	69.3	63.0	78.4	71.9	5486	.000	487 (4)

Table 6.4: Accent type prediction (plain versus nuclear) for the 4810 accented words by classifier (CART versus logistic regression (Regr)). Percentages of words correctly classified as having a plain accent (A), a nuclear accent (N) and overall accuracy (Acc) are shown for models built using different combinations of features: phrasal, including positional features (Phr incl Pos); semantic (Sem); and combination semantic, word-level acoustic and phrasal, excluding positional features (Sem+otherPh+Wd). The likelihood ratio (-2LL) and chi-squared significance tests are also given for the regression models. See text for more details.

which significantly improved the accuracy of the model were included (see Appendix E.2.2 for full list of tested and significant features).

6.2.1.2 Results and Discussion

Table 6.4 shows the performance of each classifier for each group of features, compared to a baseline, i.e. that all accents are nuclear (the most likely). As in the first experiment, accuracy figures for each outcome and overall are given for each classifier, along with the likelihood ratio of each regression model and results of chi-squared tests (the Phr and Sem feature models are compared to the baseline and the Sem+otherPh+Wd model to the Sem model).

We can see that phrasal features (Phr), including positional ones (Pos), perform very well to be able to distinguish nuclear accents from plain accents, confirming that nuclear accents do indeed tend to occur at the end of phrases. In fact, these features outperform semantic and word-level acoustic features, i.e. adding these features does not significantly improve performance over the phrase position features. This is partly what we would expect given the prosody annotation guidelines, which said that annotators should expect nuclear accents to occur toward the end of a phrase. However, annotators were given the freedom to place them earlier if they thought an earlier accent sounded more important. Further, this result is

consistent with the view that speakers will manipulate phrasal structure rather than default nuclear accent placement in order to place important information in nuclear position. It may be that our classifiers are not sensitive enough to pick up the few cases where the nuclear accent fell in non-default position because of semantic constraints.

On the other hand, our results show that semantic features (Sem) on their own do a reasonable job of distinguishing nuclear from plain accents. We will look further at which features are significant below. In the regression models, we can also see that adding general phrasal features (otherPh), e.g. the normalised mean intensity of the phrase, and word-level acoustic features (Wd), like the relativised duration of the word, has no effect on the percentage of nuclear accents correctly classified. It does lead, however, to a sizeable improvement in the percentage of plain accents identified. (Although this does not hold with the CART models). These two findings are consistent with our thesis that nuclear accent placement is strongly constrained by information and syntactic structure, while plain accents are not. The greater improvement in plain accent classification from the addition of word-level acoustic features shows our semantic features do not adequately predict these accents, while they do nuclear accents.

Table 6.5 shows the factors which significantly affect the likelihood of a nuclear, as opposed to plain, accent in the full regression model (see Appendix F.2 for a listing of all levels of all parameters). As in the first experiment, the value of Exp(B) and its significance test is shown, as well as *P diff* values.

We can see the importance of information structure: if a word is either a single kontrast (konword), or in a kontrastive NP (konnp), then it is significantly more likely to be nuclear. In fact, 60.8% of kontrastive words carry a nuclear accent, and 41.1% of words in kontrastive NPs. Interestingly, there is no effect of information status or type. We might expect that if a word was given and kontrastive, then it would be less likely to be nuclear than other kontrasts. This did not prove to be a significant effect. However, it may be that the text-based information status encoding did not capture the relevant notion of givenness (see discussion in section 2.2.2).

This table also shows us the importance of syntactic structure on accent status. Nuclear accents are more likely the further into both the clause (*propWd_cl*) and the constituent (*propWd_cns*). The proportion of constituent effect is mediated, as we might expect, by constituent type: if the constituent is a subject (subj), then a nuclear accent is less likely; whereas if it is an adjunct (adjunct) or an object (object), then a nuclear accent is more likely. Interestingly, these results look something like what we might expect given a Nuclear Stress Rule (see discussion in section 2.2.1.2). It may be that, to a much greater extent than is often claimed for English, syntactic structure itself is manipulated to keep important information in nuclear position, given that syntax so strongly constrains prosodic phrasing.

N v A	Feat	Exp(B)	P diff	Sig	Wald (df)
Increase	konnp	1.50	9.5%	.000	17.8 (1)
	konword	2.23	17.4%	.000	106.5 (1)
	numWd_cl	1.05	1.1%	.000	23.1 (1)
	propWd_cl	3.00	22.4%	.000	29.2 (1)
	propWd_cns	2.75	21.0%	.000	67.3 (1)
	adjunct by propWd_cns	1.55	10.1%	.000	32.8 (1)
	obj by propWd_cns	1.32	6.6%	.001	12.0 (1)
	numSyl_wd	1.54	9.9%	.000	95.0 (1)
	spRate_syl	1.17	3.7%	.000	30.7 (1)
	dur_relSyl	2.18	17.0%	.000	308.6 (1)
	npqrange_wd	1.18	3.9%	.000	46.6 (1)
	npmean_ph by npquan_wd	1.02	0.6%	.000	47.3 (1)
Decrease	numWd_cl by propWd_cl	0.94	-1.4%	.000	16.0 (1)
	subj by propWd_cns	0.69	-9.1%	.000	15.8 (1)
	npquan_wd	0.81	-5.3%	.000	42.0 (1)
	Constant	0.01	-	.000	203.4 (1)

Table 6.5: Factors which significantly affect accent type prediction, i.e the likelihood of a nuclear, not plain, accent in the full regression model excluding phrasal position factors (i.e. Sem+otherPh+Wd, see Appendix E for description of features). The odds ratio for each feature in the model (Exp(B)), as well its significance (Sig) using the Wald statistic is given, along with the percentage change (P diff) in the likelihood of a nuclear accent with the presence of that level of that variable (or a one unit increase for continuous variables).

Finally, we can see that the likelihood of a nuclear accent increases with the duration of the word (*dur_relSyl*), the number of syllables in the word (*numSyl_wd*) and the normalised pitch quantile range of the word (*npqrange_wd*). This shows that nuclear accents do tend to be more acoustically prominent than plain accents. However, as we saw, addition of these features has a greater effect on the percentage of plain accents recognised than nuclear accents. Therefore, this finding actually serves to show again that, in relation to nuclear accents, plain accents can only be predicted by their acoustic properties, not their semantic ones.

Overall, this experiment shows that, consistent with our claims, nuclear accents can be

distinguished from plain accents by their structural properties. Phrasal properties are most effective, but kontrast and syntax structure are also reasonable predictors. Word-level acoustic features only help in the identification of plain accents, in line with the view that these accents cannot be readily defined in terms of their semantic properties.

6.2.2 A2: Nuclear Accents are Meaningful

The second experiment compares two sets of models. The first try to distinguish plain accented from unaccented words, the second set nuclear accented from unaccented words.

6.2.2.1 Aim and Method

Our general claim and the specific hypothesis being tested in this experiment are:

- **General Claim:** Kontrast aligns with the most prominent prosodic node in its scope. When there is only one kontrast in a phrase, this is the nuclear accent. Therefore we expect most other accents to appear only as required by the metrical structure and not to be directly meaningful.
- **Hypothesis:** Nuclear accented words can be reliably distinguished from unaccented words by semantic/syntactic and phrase-level features. Other accents can only be reliably predicted by a combination of these features with word-level acoustic features.

The method in this experiment was to build two sets of prediction models, and compare the results of each using the two classifiers as before. The first set of models predicted the probability of a word having an accent, with nuclear accented words excluded from the data set, thereby distinguishing plain accented from unaccented words. The second set of models predicted the probability of a word having a nuclear accent, with plain accented words excluded, thereby distinguishing nuclear accented from unaccented words. The same data set was used as in the first two experiments, with the noted exclusions. There were 6140 words in the plain/unaccented comparison, and 6880 in the nuclear/unaccented set. The same feature set was used as in the last experiment, except that all phrasal features were grouped, for a total of 18 phrasal features, 16 semantic/syntactic features and 8 word-level acoustic features. Feature reduction was carried out as in the last experiments, so that only significant features were included in reported models (see Appendices E.2.3 and E.2.4 for full lists of tested and significant features).

+A v –A	CART Classifier			Regr Classifier			Regr Model Fit		
	–A	+A	Acc	–A	+A	Acc	-2LL	Sig	χ^2 (df)
Baseline	100	0	68.2	100	0	68.2	8724	-	-
Phr	83.5	50.0	72.4	90.9	44.0	76.0	7018	.000	1707 (10)
Sem	85.8	34.7	68.8	90.9	31.8	72.1	7792	.000	910 (29)
Phr + Sem	84.5	52.2	73.8	89.9	50.9	77.5	6525	.000	428 (17)
Phr+Sem+Wd	86.2	56.4	76.4	89.4	62.9	80.6	5432	.000	811 (3)
N v –A	**–A**	**N**	**Acc**	**–A**	**N**	**Acc**	**-2LL**	**Sig**	χ^2 **(df)**
Baseline	100	0	60.6	100	0	60.6	10534	-	-
Phr	86.1	73.5	81.0	87.4	76.3	83.0	5939	.000	4595 (12)
Sem	81.9	70.5	77.3	82.3	70.7	77.7	7622	.000	2874 (23)
Phr + Sem	87.9	78.3	84.0	90.0	79.7	85.9	5116	.000	823 (5)
Phr+Sem+Wd	88.0	79.3	84.5	91.1	84.0	88.2	4014	.000	766 (4)

Table 6.6: Comparison of plain accent versus no accent (6140 words), and nuclear accent versus no accent (6880 words) prediction by classifier (CART versus logistic regression (Regr)). Percentages of words correctly classified as having no accent (–A), having either a plain or a nuclear accent (+A and N) and overall accuracy (Acc) are shown for models built using different combinations of features: phrasal (Phr); semantic (Sem); combination phrasal and semantic (Phr+Sem); and phrasal, semantic combined with word-level acoustic (Phr+Sem+Wd). The likelihood ratio (-2LL) and chi-squared significance tests are also given for the regression models. See text for more details.

6.2.2.2 Results and Discussion

Table 6.6 shows the performance of each classifier on each group of variables for each set of models using the same format as before, compared to the baseline where all words are unaccented. We can see that in all cases the combined feature models lead to improvement over the previous model both in terms of prediction accuracy, and, in the regression case, significant reduction in the likelihood ratio (for both sets the chi-squared test compares the Phr and Sem regression models to the baseline; the Phr+Sem model to the Phr model; and the Phr+Sem+Wd model to the Phr+Sem model).

We can immediately see that prediction accuracy for nuclear accents is significantly better overall than for plain accents. This concurs with our general claim that the occurrence

of nuclear accents is more predictable than the occurrence of other accents. More crucially, we can see that while performance on the prediction of nuclear accents using only semantic (Sem) or phrasal (Phr) features is reasonable (between 70.5-76.3% accents correctly identified), for plain accents prediction using only these features is poor (31.8-50.0%). This is consistent with our contention that nuclear accent placement is constrained by the types of information, syntactic and phrasal structure features we are looking at here, whereas other accents are much less so. We can see this particularly with the contribution of semantic features. While the percentage of nuclear accents correctly identified using only semantic features is 70.5-70.7%, for plain accents this is only 31.8-34.7%, i.e. considerably less than chance. This is not simply a matter of the reliability with which plain accents were annotated compared to nuclear accents; as there was no difference in inter-annotator agreement over all accent types, compared to the presence/absence of an accent (see section 5.4.3).

Lastly, over both the classifiers the addition of word-level acoustic features leads to a much more substantial increase in accent recognition for plain accents than for nuclear accents. This increase is greater for the regression classifier, whose performance generally improves much more with the addition of semantic and word-level features. It is not clear why this is. It could be because regression results are slightly inflated, although the regression classifier does not seem to have performed substantially better over the results as a whole. The important point, however, is that in both cases word-level features lead to a greater improvement in plain accent than nuclear accent recognition. We know that accented words, of whatever type, are likely to be more acoustically prominent than unaccented words. Therefore this result is further evidence for our contention that nuclear accent placement is constrained by the type of semantic and phrasal features we are looking at here; whereas plain accent placement is much less so, and so these accents can only be predicted by their acoustic features.

Tables 6.7 and 6.8 show the factors which significantly affect the likelihood of a plain accent and a nuclear accent respectively in the final regression models, using the same format as for the last two experiments (see Appendices F.3 and F.4 for a listing of all levels of all parameters in each model). A comparison of the significant factors in each model will help us to further illustrate the claims just made.

In both cases we can see that if the word is a single kontrast (konword), this substantially increases the probability of an accent. However, for nuclear accents this increase is greater and there is also a substantial effect if the word is in a kontrastive NP (konnp). This is what we would expect, since an entire kontrastive NP would be more likely to head its own phrase, whereas a single kontrastive word may only be accented to mark its status. In both models, the interaction with the number of distinct kontrasts in the phrase (*numKon_ph*) decreases the

+A v –A	Feat	Exp(B)	P diff	Sig	Wald (df)
Increase	konword	4.33	35.1%	.000	45.5 (1)
	adjunct	1.58	10.6%	.013	6.1 (1)
	NN	1.48	9.1%	.039	4.3 (1)
	DT	1.52	9.7%	.038	4.3 (1)
	numSyl_wd	2.90	25.7%	.000	222.3 (1)
	t_ph	1.23	4.6%	.000	29.2 (1)
	brk	1.61	11.0%	.006	7.7 (1)
	accq_dist	1.77	13.4%	.000	119.4 (1)
	accsPh_exc by numWd_ph	1.19	3.9%	.000	34.5 (1)
	npquan_wd	1.43	8.2%	.000	257.8 (1)
	nimean_wd	1.63	11.4%	.000	178.7 (1)
	dur_relSyl	2.44	21.4%	.000	258.5 (1)
Decrease	PR	0.58	-10.6%	.007	7.4 (1)
	konword by numKon_ph	0.51	-12.6%	.001	11.8 (1)
	adjunct by propWd_cns	0.56	-11.1%	.013	6.2 (1)
	obj by propWd_cns	0.57	-10.8%	.012	6.2 (1)
	numWd_ph	0.91	-2.0%	.001	10.2 (1)
	npmean_ph	0.77	-5.3%	.000	117.5 (1)
	propWd_ph	0.18	-24.1%	.000	21.2 (1)
	accsPh_exc	0.07	-28.7%	.000	139.6 (1)
	numWd_ph by t_ph	0.99	-0.3%	.000	12.5 (1)
	Constant	0.00	-	.000	214.7 (1)

Table 6.7: Factors which significantly affect the likelihood of a plain accent, as opposed to no accent, in the full regression model (i.e. Phr+Sem+Wd, see Appendix E for description of features). The odds ratio for each feature in the model (Exp(B)), as well its significance (Sig) using the Wald statistic is given, along with the percentage change (P diff) in the likelihood of a plain accent with the presence of that level of that variable (or a one unit increase for continuous variables).

likelihood of an accent, but more so with nuclear accents. This follows on from the above: if there is more than one kontrast in a phrase, only one can align with the nuclear accent. Therefore the probability of each kontrastive word carrying a nuclear accent decreases.

N v –A	Feat	Exp(B)	P diff	Sig	Wald (df)
Increase	konnp	3.82	31.9%	.000	26.8 (1)
	konword	21.45	53.9%	.000	170.9 (1)
	accq_dist	1.88	15.6%	.000	123.3 (1)
	accsPh_exc by numWd_ph	1.38	7.9%	.000	130.2 (1)
	numWd_ph by propPho_ph	3.29	28.8%	.000	101.3 (1)
	spRate_syl by t_ph	1.08	1.9%	.000	262.7 (1)
	npquan_wd	1.57	11.2%	.000	221.3 (1)
	nimean_wd	1.56	11.0%	.000	94.5 (1)
	dur_relSyl	2.22	19.6%	.000	301.9 (1)
Decrease	numKon_ph	0.59	-11.7%	.000	29.2 (1)
	konword by numKon_ph	0.56	-12.6%	.001	10.2 (1)
	posWd_ph	0.37	-20.2%	.000	71.0 (1)
	numWd_ph	0.55	-13.0%	.000	197.6 (1)
	accsPh_exc	0.07	-35.2%	.000	158.0 (1)
	brk	0.37	-20.1%	.000	57.2 (1)
	numWd_ph by posSyl_ph	0.95	-1.2%	.000	75.0 (1)
	accsPh_exc by npmean_ph	0.88	-3.1%	.000	51.9 (1)
	npmin_wd	0.80	-5.3%	.000	62.8 (1)
	Constant	0.00	-	.000	152.6 (1)

Table 6.8: Factors which significantly affect the likelihood of a nuclear accent, as opposed to no accent, in the full regression model (i.e. Phr+Sem+Wd, see Appendix E for description of features). The odds ratio for each feature in the model (Exp(B)), as well its significance (Sig) using the Wald statistic is given, along with the percentage change (P diff) in the likelihood of a nuclear accent with the presence of that level of that variable (or a one unit increase for continuous variables).

In this comparison, there are no significant syntactic features in the nuclear accent model. But we can see that the likelihood of a plain accent increases if the word is in an adjunct (adjunct) or a noun (NN), and decreases if it is a pronoun (PR). This evidence could lead us to reassess our analysis of the effect of syntactic features in the last experiment. Rather than certain types of constituent making a nuclear accent more likely, these features could have been serving to positively identify plain accents (like with the acoustic features). That is,

certain types of constituents/part-of-speech types are inherently more 'prominence-lending', and are therefore more likely to surface accented, whatever their information status. We can see that the likelihood of a plain accent in an object decreases as the proportion of the object increases (obj by *propWd_cns*). The end of the object would be the 'default' nuclear accent position, which could explain this result. If this is so, it would also support the suggestion we made in the discussion of the last experiment that syntactic structure may be manipulated to put kontrastive elements in nuclear position, rather than syntax acting as a direct constraint on nuclear accent placement.

Once more, we can see the effect of the constraint placing nuclear accents towards the end of phrases: a nuclear accent is much more likely as the proportion of words in the phrase increases (*numWd_ph* by *propPho_ph*), and a plain accent much less likely (*propWd_ph*). Finally, there are a number of features which we would expect to predict plain accents which actually have a similar effect on the likelihood of plain and nuclear accents. The probability of both increases as the number of words since the last accent increases (*accq_dist*): this structural requirement holds on both types. It also increases as the normalised quantile pitch (*npquan_wd*), normalised mean intensity (*nimean_wd*), and relative duration (*dur_relSyl*) increase. As we discussed above, this is not surprising since we know that both types of accent are more acoustically prominent than unaccented words. However, again these features substantially improve recognition for plain accents but not nuclear accents. Therefore, we can say that these features add very little new information to the nuclear accent prediction model, that is not provided by the semantic and phrasal features; whereas they do for the plain accent model.

The evidence from this experiment seems to support quite clearly our contention that nuclear accents are 'meaningful', i.e. they can be successfully predicted by information and syntactic structure features; where plain accents are much less so. Plain accent recognition is much poorer overall, and reasonable recognition levels can only be achieved by including word-level acoustic features. This is consistent with our claim that most of these accents do not directly reflect information structure. Their placement is either much more arbitrary, or is determined by rhythm, constituent type or other constraints not well captured in the feature set here.

6.2.3 A3: Accent and Nuclear Accent Prediction

The final accent experiment seeks to consolidate the findings of the first two by presenting overall accent prediction models including all of the data. Firstly, we report models which predict whether a word is accented or not, including both accent types. Secondly, we report models which distinguish unaccented, plain and nuclear accented words.

6.2.3.1 Aim and Method

Our general claim and the specific hypotheses being tested in this experiment are:

- **General Claim:** Prosodic prominence patterns result from the interaction of the information and syntactic properties of each word, constrained by prosodic phrasing and metrical structure. In particular, phrasal structure constrains the perception of nuclear prominence.
- **Hypothesis 1:** The most effective model for predicting accents and nuclear accents includes information structure, syntax, phrasal-level phonetic and word-level acoustic features.
- **Hypothesis 2:** Nuclear accent and accent type prediction is substantially improved by prosodic phrasal features.

In this experiment we report two sets of prediction models. The first set predict the probability of each word having an accent (either plain or nuclear), using CART and logistic regression classifiers. The second set predicts the accent group of the word, i.e. unaccented, plain or nuclear accented, using CART and multinomial logisitic regression. This is a form of logistic regression allowing more than two levels in the dependent variable (see summary in section 6.0.2). The data set is the same as in the earlier experiments, including all 8915 words. The same feature set was used as in the last experiment, except that for the purposes of testing the second hypothesis, in the accent group set phrasal features were divided into 'phrase proportion features' and other phrasal features. Phrase proportion features are those which explicitly refer to where the current word is in a completed phrase. These features most directly fall out from the idea that nuclear accent placement is strongly constrained by phrasal structure, and so will be used to test this claim. As before, only significant features within each feature group were included in the reported models (see Appendices E.2.5 and E.2.6 for full lists of tested and significant features).

6.2.3.2 Results and Discussion

Table 6.9 shows the recognition accuracy for binary accent prediction for each outcome by classifier and feature type, as well as the likelihood ratio for the regression models as before. We can see that, for the regression models, each model does significantly reduce the unexplained variance compared to the last (the chi-squared tests compare the Phr and Sem models to the baseline, the Phr+Sem model to the Phr model, and the Phr+Sem+Wd model to the Phr+Sem model). Overall, the regression classifier was better able to model this data than CART. In particular, while the addition of semantic features (Sem) led to a small, but

A/N v –A	CART Classifier			Regr Classifier			Regr Model Fit		
	–A	+A	Acc	–A	+A	Acc	-2LL	Sig	χ^2 (df)
Baseline	0	100	52.8	0	100	52.8	13931	-	-
Phr	71.8	74.3	73.1	76.9	74.3	75.5	10000	.000	3931 (13)
Sem	67.9	71.3	69.7	74.0	68.5	71.1	11442	.000	2470 (35)
Phr+Sem	76.0	75.5	75.7	78.9	77.5	78.2	9036	.000	862 (22)
Phr+Sem+Wd	75.5	80.5	78.2	81.5	83.1	82.4	7366	.000	1466 (3)

Table 6.9: Accent prediction for all 8915 words by classifier (CART versus logistic regression (Regr)). Percentages of words correctly classified as not having an accent (–A), having either a plain or nuclear accent (+A) and overall accuracy (Acc) are shown for models built using different combinations of features: phrasal (Phr); semantic (Sem); combination phrasal and semantic (Phr+Sem); and phrasal, semantic combined with word-level acoustic (Phr+Sem+Wd). The likelihood ratio (-2LL) and chi-squared significance tests are also given for the regression models. See text for more details.

noticeable increase in the number of accents recognised with the regression model, it led to only a very small increase with the CART model. This could be due to the nature of the classifiers: firstly the regression model can explicitly test for the interaction of different features, such as kontrast status and part-of-speech type; whereas in the CART model these interactions have to fall out from the tree structure. Secondly, the regression model captures overall tendencies better, since every feature has a weight on each data point, while CART captures local effects better, within sub-branches of the tree. In this case, overall tendencies, such as the normalised mean pitch of the phrase, and the word mean pitch, seem to be important.

Table 6.10 shows the features that were significant in the final model for each classifier. As can be seen, there are very few semantic features, primarily kontrast status (*kon_stat*) and part-of-speech group (*POS_gp*). Given the results in the last experiment, this result is probably indicative of the effectiveness of kontrast in predicting nuclear accent status, and the ineffectuality of any semantic features in predicting plain accents. Therefore, overall kontrast and part-of-speech are the best semantic predictors of accenting; not because they actually predict all accents well, but because they predict nuclear accents well, and other accents cannot be effectively identified by other semantic features included here. The regression classifier uses a lot more phrasal features than CART, which may help explain why its

A/N v –A	CART Classifier	Regr Classifier
Phr	numSyl_wd, numSyl_ph, t_ph, propWd_ph, nimean_ph, spRate_syl, accPh_exc	numSyl_wd, is_break, numWd_ph, t_ph, npmean_ph, propWd_ph, propPho_ph, accq_dist, accPh_exc, accPh_exc*numWd_ph, spRate_syl*t_ph, numWd_ph*t_ph
Sem	kon_stat, clause_type, POS_gp	kon_stat, kon_stat*POS_gp
Wd	npmean_wd, dur_relSyl, nprange_wd	npquan_wd, nimean_wd, dur_relSyl

Table 6.10: Significant features in the final accent prediction models (Phr+Sem+Wd) by classifier (CART and logistic regression (Regr)). Features are grouped by type: phrasal (Phr), semantic (Sem) and word-level acoustic (Wd) (see Appendix E for description of features).

performance is better. Again, these may have been more effective in the regression model because they are overall tendencies, rather than localised effects.

One reason for including this set of models was to enable these results to be compared to previous reported accent prediction studies. As we saw in section 6.0.1, current state-of-the-art accuracy is around 84.0-88.3%. Our results are lower than this, with 82.4% for the regression model and 78.2% for CART. However, these reported studies used read, monologue or radio news speech, which tends to have more regular prosodic patterns than spontaneous speech. Our model also covers many more speakers than most of those studies. The fact that these results are at all comparable, using just two semantic features and a variety of phrasal constraints, could be taken as evidence in the debate set out in section 2.2.2.2: whether the part-of-speech, and other low-level features used in those studies are successful in predicting accents because prominence is directly related to these low level features; or because these features correlate with higher level information structure, which in turn constrains prominence. The results of the experiments presented so far suggest that it is not enough to simply look at 'accents'. Nuclear accents are strongly constrained by kontrast status, within a phrasal structure that is constrained by syntax. Plain accents, on the other hand, may in part be directly predicted by these low-level features.

Table 6.11 shows accent group classification results by classifier for each feature type. As before, recognition rates by accent group are given, along with overall accuracy. This is

N v A v –A	CART Classifier				Regr Classifier				Regr Model Fit		
	–A	A	N	Acc	–A	A	N	Acc	-2LL	Sig	χ^2 (df)
Baseline	100	0	0	47.2	100	0	0	47.2	19252	-	-
Phr	77.5	40.5	65.8	65.4	82.9	29.0	71.9	61.5	16631	.000	2621 (22)
Sem	77.8	17.7	61.0	58.8	79.9	11.0	68.1	58.4	12425	.000	2142 (56)
Phr+Sem	79.5	38.7	71.8	67.8	84.8	35.0	75.0	64.3	15644	.000	3446 (52)
P+S+Wd	78.6	44.1	70.5	68.2	85.8	47.2	77.7	67.7	13822	.000	4317 (52)
PS-Phpos	79.4	26.3	62.7	62.1	85.3	19.1	66.6	59.0	17232	.000	1798 (42)

Table 6.11: Accent group prediction for all 8915 words by classifier (CART versus multinomial logistic regression (Regr)). Percentages of words correctly classified as having no accent (–A), having a plain accent (A), having a nuclear accent (N) and overall accuracy (Acc) are shown for models built using different combinations of features: phrasal (Phr); semantic (Sem); combination phrasal and semantic (Phr+Sem); phrasal, semantic combined with word-level acoustic (P+S+Wd); and the phrasal/semantic model excluding phrase positional features (PS-Phpos). The likelihood ratio (-2LL) and chi-squared significance tests are also given for the regression models. See text for more details.

compared to a baseline where all words are unaccented. We can see that, overall, phrasal features (Phr) performed better than semantic features (Sem). This is particularly true for plain accent recognition, where percentage accuracy is well below chance (33%) using semantic features only. This nicely confirms the claims in the last two studies. Firstly, we showed that nuclear and plain accents can best be separated in terms of their phrasal structure properties. Here, we show again that nuclear accents can be recognised reasonably well using only phrasal features; and secondly, that plain accent recognition does improve using these features, showing their appearance is at least in part constrained by phrasal properties in the rest of the phrase. Secondly, we showed that nuclear accents are more 'meaningful' than plain accents (in terms of our semantic/syntactic features), which is clearly confirmed here. Further, the addition of semantic features to the phrasal feature model (Phr+Sem) leads to a sizeable increase in nuclear accent recognition; but no increase in plain accent recognition in the CART model (though it does in the regression model). This shows again that these semantic and phrasal features are adding different, and real, constraints to nuclear accent placement; but that the semantic constraints on plain accent placement are much weaker. Finally, we can see once more that the addition of word-level acoustic features substan-

tially improves plain accent recognition, but leads to only a small increase in nuclear accent recognition. This further supports our contention that, since both types of accent are more acoustically prominent than unaccented words; nuclear accents are already well predicted by semantic and phrasal features, while plain accents are not.

In relation to the second hypothesis, we can see quite clearly that phrasal position features make a substantial difference to the effectiveness of the model. Recognition rates for both nuclear and plain accents drop when these features are excluded. If we compare the reduction in performance here to the reduction in performance resulting from the exclusion of accentual features in the phrase prediction model (see Table 6.1), we can see the effect is much more dramatic. This supports our argument there that phrase placement 'comes first', in the sense that phrasal structure strongly constrains prominence structure (as we show here for both nuclear and plain accents), rather than the other way around.

Tables 6.12 and 6.13 show the factors which significantly affected the probability of a plain accent and a nuclear accent respectively in the full regression model, using the same format as before (see Appendix F.5 for a listing of all levels of all parameters in the model). Comparison of the significant factors in these tables confirms the trends noted in the past two experiments and our claims about these made then. We can see once more that the word being a kontrast (konword) makes both types of accent more likely, but this effect is stronger with nuclear accents, and holds for words in kontrastive NPs (konnp) as well. Again, no syntactic features are significant in predicting nuclear accents. However, a plain accent is less likely if the word is an adverb (RB), pronoun (PR), or a verb (VB). By extension, this would mean the other part-of-speech types are more 'accentable'. The fact this is not a significant feature for nuclear accents could indicate that kontrast status overrides any inherent prominence because of part-of-speech type.

Once more we see the effect of phrase position features. The proportion of phones in the phrase (*propPho_ph*) makes a plain accent considerably less likely. As in the previous studies, word-level acoustic features, especially quantile pitch (*npquan_wd*), mean intensity (*nimean_wd*) and relative duration (*dur_relSyl*), make both plain and nuclear accents substantially more likely. Again, since these features only substantially improve plain accent recognition, this suggests that the other features do not otherwise predict these accents well, whereas they do nuclear accents.

Overall, these experiments uphold the predictions of our theory on the relationship between prominence, information structure and prosodic structure set out in Chapter 3. There we claimed that accents were a manifestation of prominent elements in a metrical prosodic structure, rather than being independent entities. We predicted from this that the manifestation of accents and accent type (plain/nuclear) would be strongly constrained by phrasal

A v –A/N	Feat	Exp(B)	P diff	Sig	Wald (df)
	Intercept	0.00	-	.000	211.8 (1)
Increase	konword	2.92	23.5%	.000	28.5 (1)
	backgd * numKon_ph	1.18	3.0%	.044	4.1 (1)
	numSyl_wd	2.28	17.5%	.000	213.1 (1)
	accq_dist	1.34	5.6%	.000	63.7 (1)
	propPho_ph * numPho_ph	1.08	1.5%	.001	11.7 (1)
	spRate_syl * t_ph	1.02	0.3%	.003	9.1 (1)
	npquan_wd	1.27	4.5%	.000	188.1 (1)
	nimean_wd	1.41	6.6%	.000	128.3 (1)
	dur_relSyl	2.01	14.5%	.000	320.4 (1)
Decrease	RB	0.75	-4.6%	.029	4.8 (1)
	PR	0.56	-8.6%	.000	21.2 (1)
	VB	0.73	-5.1%	.007	7.4 (1)
	konword * numKon_ph	0.72	-5.2%	.032	4.6 (1)
	accsPh_exc	0.33	-14.0%	.000	101.9 (1)
	npmean_ph	0.84	-3.0%	.000	85.7 (1)
	propPho_ph	0.15	-18.7%	.000	32.1 (1)
	numPho_ph	0.96	-0.7%	.000	27.6 (1)

Table 6.12: Factors which significantly affected the likelihood of a plain accent in the full regression model (i.e. P+S+Wd, see Appendix E for description of features). The odds ratio for each feature in the model (Exp(B)), as well its significance (Sig) using the Wald statistic is given, along with the percentage change (P diff) in the likelihood of a plain accent with the presence of that level of that variable (or a one unit increase for continuous variables).

structure. Here we have shown that this prediction bears out: firstly, nuclear accents can be most effectively distinguished from plain accents by phrasal position features; secondly, plain accent prediction is very poor if phrasal features (such as position and distance to the last accent) are not included; lastly, accent group prediction is substantially worse without the inclusion of phrase proportion features. We further claimed a strong relationship between this prosodic structure and information structure, namely that kontrastive elements want to align with nuclear accents within a phrase structure constrained by syntax. Other accents

N v A/–A	Feat	Exp(B)	P diff	Sig	Wald (df)
	Intercept	0.00	-	.000	260.8 (1)
Increase	konnp	1.68	12.1%	.014	6.0 (1)
	konword	5.24	39.2%	.000	81.1 (1)
	brk	1.82	14.1%	.000	23.8 (1)
	numSyl_wd	2.38	20.7%	.000	222.3 (1)
	accq_dist	1.34	6.7%	.000	62.3 (1)
	propPho_ph * numPho_ph	1.12	2.5%	.000	20.9 (1)
	accsPh_exc * numWd_ph	1.07	1.6%	.000	33.7 (1)
	spRate_syl * t_ph	1.06	1.2%	.000	84.5 (1)
	npquan_wd	1.36	7.0%	.000	258.2 (1)
	nimean_wd	1.50	9.3%	.000	138.5 (1)
	dur_relSyl	2.02	16.6%	.000	333.1 (1)
	npqrange_wd	1.04	0.8%	.011	6.4 (1)
Decrease	konword * numKon_ph	0.69	-7.5%	.004	8.5 (1)
	posSyl_ph	0.61	-9.5%	.000	47.2 (1)
	accsPh_exc	0.24	-21.4%	.000	196.5 (1)
	npmean_ph	0.77	-5.3%	.000	146.4 (1)
	numPho_ph	0.89	-2.5%	.000	94.5 (1)

Table 6.13: Factors which significantly affected the likelihood of a nuclear accent in the full regression model (i.e. P+S+Wd, see Appendix E for description of features). The odds ratio for each feature in the model (Exp(B)), as well its significance (Sig) using the Wald statistic is given, along with the percentage change (P diff) in the likelihood of a nuclear accent with the presence of that level of that variable (or a one unit increase for continuous variables).

appear for rhythmical reasons within this structure. They are less directly meaningful, although certain syntactic elements may be more likely to be accented than others. In the studies presented this claim seems to be confirmed: kontrast status is a strong predictor of nuclear accents; other accents are not well predicted by semantic features, although certain types of constituent or part-of-speech type may be inherently more 'prominence-lending'.

6.3 Kontrast and Prominence

In the experiments so far we have seen that there is a general correspondence between nuclear accenting and kontrast, i.e. kontrasts within information units tend to align with prominent positions in the prosodic structure. However, it remains true that not all kontrasts are marked by nuclear accents, and not all nuclear accents are kontrastive. Therefore, we need to look more closely at what factors make a kontrastive interpretation more likely, along with structural prosodic prominence.

Up to this point, we have been primarily talking about prosodic prominence in terms of its structural properties. However, this structure is of course in part conveyed by the acoustic properties of elements within it. In section 3.1.2 we laid out the properties of a basic prosodic phrase: with pre-nuclear accents being of roughly equal acoustic prominence to nuclear accents; and post-nuclear accents, if they appear at all, being much less prominent. However, as discussed in section 3.1.4, these same acoustic properties are also manipulated to raise the emphasis of a single word, or the entire phrase. It is not clear exactly how this interacts with the perception of structural prominence, as a very high early accent can override the expectation of a nuclear accent late in the phrase, but not necessarily so. It is also not clear if 'raising for emphasis' on a single word is gradient or categorical (or both), i.e. whether there is a real perceptual category of 'emphatic accents' (see section 3.2.4). In section 3.2.2, we claimed that both types of prosodic prominence are manipulated to convey kontrast and information structure. That is, an element is likely to be interpreted as kontrastive if it is in a position which is structurally prominent, but that this is tempered by the likelihood that it would be in such a position anyway. So a pronoun that is accented would be more likely to be kontrastive than an accented noun because it would usually be deaccented. The likelihood of a kontrast also increases the more acoustically prominent it is, compared to its expected prominence. In other words, a kontrastive interpretation is more likely if an element is more prominent (both structurally and acoustically) than expected given its semantic, syntactic and discourse features; as well, of course, as the plausibility of a kontrastive interpretation in the context. This is particularly true of *restricted* kontrast readings which, as we suggested in section 2.2.1.3, become more likely the more prominent an element is.

In the following experiments we wish to test some of these claims on our corpus. In the first set of experiments, we will seek to show that kontrast status can be predicted reasonably well from either structural prominence or semantic/syntactic features, but that this improves if we consider the likelihood of the acoustic prominence of the word given those features. In the second set, we look more closely at the acoustic features of different accents. We show that, overall, accent types (pre-, post-, nuclear) do indeed have distinct acoustic profiles; and, secondly, that this is manipulated to show the kontrast status of the accent. We will consider

what this means in terms of kontrast and *restricted* kontrast interpretations.

6.3.1 K1: Kontrasts are More Prominent than Expected

The first experiment therefore built models to predict whether a word is kontrastive or not. We test the relative effectiveness of structural prosodic prominence, semantic/syntactic and word-level acoustic features.

6.3.1.1 Aim and Method

Our general claim and the specific hypothesis being tested in this experiment are:

- **General Claim:** The kontrast status of an element is determined by its prominence in context. That is, its structural prosodic prominence along with the likelihood that it would be both as structurally and acoustically prominent as it is given its semantic and syntactic properties.
- **Hypothesis:** A word is likely to be kontrastive if it is inherently prosodically, semantically or syntactically strong, or if it is more prominent than expected given its information status and syntactic features.

The method used in this experiment was to build models to predict whether a word is kontrastive or not. Kontrast could be either at the NP or the word level. In these experiments, only the regression classifier was used, as we were interested in gauging the relative effectiveness of different types of features, and particularly their interaction, something that it is harder to control with CART. Further, we did not seek to compare the results to related work, as kontrast (as we define it), is a new feature and so there are no analogous studies. In the first set of models, all words were included, using the same data set as for the earlier studies, a total of 9289 words.[1] The second set of models looked specifically at kontrast prediction on nuclear accented words, therefore plain and unaccented words were excluded, leaving a total of 2927 words.

The feature set was reduced compared to the previous experiments, based on the features which had proved significant in the accent group prediction model in A3. For the first set of models, we were particularly interested in the effect of accent group (nuclear, plain, unaccented), so included only a limited number of other phrasal features. We only included *propSyl_ph* (position of the current syllable relative to the number of syllables in the phrase) as a phrasal position feature as exploratory tests showed this performed similarly to a fuller

[1] There were slightly more words that in the previous experiments because some features with missing values were not included.

range of position features and it was easier to interpret in terms of interaction with semantic features. Other phrasal features that had been significant in the accent group model were also tested (see Appendix E.2.6), only *numSyl_wd* (number of syllables in the word) and *spRate_syl* (speech rate in syllables per second) were significant. Semantic/syntactic features that were significant in the final accent group model were tested (excluding kontrast features). We also included information status as a feature. This was not significant in the accent group model, but this could have been because kontrast captured the same information better. The same set of phrasal and semantic/syntactic features were used in the kontrast prediction model on nuclear accents (obviously excluding accent group). In both cases, features which were not significant were excluded.

Finally, a single measure of *prominence* was included in both sets of models to gauge the effect of increasing the acoustic prominence of the word. This proved more consistent than trying to include the interaction of the different acoustic correlates of prominence separately. However, this was a fairly rough measure, devised by approximating the relative contribution of the different acoustic variables that were significant in the accent group model; while trying to roughly equalise their means and standard deviations and achieve a range similar to that of the original measures. The final measure was computed as follows:

$$prom = ((2 * dur\_relSyl) + nqrange\_wd + npquan\_wd + (nimean\_wd - 5))/10 \qquad (6.5)$$

6.3.1.2 Results and Discussion

Table 6.14 shows the performance of the regression model on kontrast prediction for each group of features, along with the likelihood scores for each model and chi-squared tests (the Phr and Sem model are compared their baselines, and the S+Ph+Prom models to their respective Sem models). The results show that broadly, our hypothesis is confirmed. The simple accentual/phrasal model (Phr) and our semantic/syntactic model (Sem) perform comparably to classify words as kontrastive. This shows once more the strong correspondence between prosodically strong positions and kontrast on the one hand, and semantically/syntactically strong constituents and kontrast on the other. Further, the interaction of prominence and semantic features (S+Ph+Prom) leads to a substantial improvement in the number of kontrasts identified. We will see more clearly below how this confirms our hypothesis that a kontrastive interpretation is more likely the more prominent an element is than expected.

The second set of models look more particularly at what leads to a nuclear accent being interpreted as kontrastive. We can see that the mere fact that it is nuclear strongly biases a kontrast. Further, the combination of semantic/syntactic features (Sem) we include lead to extremely good identification of which nuclear accents are kontrastive. Our models performs

+K v –K		Regr Classifier			Regr Model Fit		
		–K	+K	Acc	-2LL	Sig	χ^2 (df)
All Words	Baseline	100	0	60.0	13650	-	-
	Phr	80.9	60.1	72.6	11256	.000	2394 (5)
	Sem	81.4	65.1	74.8	9851	.000	2705 (14)
	S+Ph+Prom	80.0	73.6	77.4	8816	.000	1034 (31)
Nuclear	Baseline	0	100	69.2	3614	-	-
	Sem	28.2	92.3	72.5	3277	.000	337 (8)
	S+Ph+Prom	38.1	90.7	74.5	3075	.000	202 (11)

Table 6.14: Kontrast prediction over all 9289 words, and for nuclear accented words only (2927), using the logistic regression classifier. Percentages of words correctly classified as being not kontrastive (–K), kontrastive (+K) and overall accuracy (Acc) are shown for models built using different combinations of features: phrasal (Phr); semantic (Sem); and combination phrasal, semantic with the acoustic prominence measure (S+Ph+Prom). The likelihood ratio (-2LL) and chi-squared significance tests are also given for the regression models. See text for more details.

very badly, however, at determining which nuclear accents are not kontrastive. The interaction with phrasal and prominence features (S+Ph+Prom) helps considerably, but identification of backgrounded nuclear accents is still poor. This could be due in part to our kontrast coding scheme, which required kontrasts to fall into a number of categories, and may therefore have led to a number of 'false-negative' background classifications, as we discussed in the last chapter. It may also be due to the rough nature of the prominence measure. However, the more likely explanation arises from two factors that are very hard to capture in this type of model. Firstly, information structure interpretation arises not only from the type of variables discussed here, but the plausibility of any given information structure reading in the context. It is very difficult to find a good measure of plausibility. Secondly, as we discussed in section 3.2.3, the association between nuclear accenting and kontrast may not always be at the lowest level of prosodic phrasing, but may be interpreted over several phrases in a recursive structure. We will return to this point in the final experiment.

+K v –K	Feat	Exp(B)	P diff	Sig	Wald (df)
Increase	med	2.26	20.1%	.000	32.4 (1)
	JJ	8.31	44.7%	.000	24.9 (1)
	VB	2.66	24.0%	.022	5.3 (1)
	NN	6.70	41.7%	.000	20.6 (1)
	DT	2.47	22.2%	.035	4.5 (1)
	propWd_cl	3.10	27.4%	.000	117.3 (1)
	new by propSyl_ph	2.31	20.7%	.000	13.8 (1)
	head_cns by propSyl_ph	2.72	24.5%	.000	25.5 (1)
	accq by JJ	10.44	47.4%	.019	5.5 (1)
	accq by PR	9.67	46.6%	.012	6.3 (1)
	nuc by PR	21.36	53.4%	.001	11.1 (1)
	nuc by VB	6.53	41.3%	.001	10.4 (1)
	adjunct by prom_wd	1.19	4.3%	.015	5.9 (1)
	obj by prom_wd	1.16	3.5%	.024	5.1 (1)
	nuc by RB by prom_wd	3.35	29.0%	.019	5.5 (1)
	nuc by NN by prom_wd	1.89	15.8%	.015	5.9 (1)
Decrease	head_cns	0.29	-23.6%	.000	64.5 (1)
	adjunct by propWd_cns	0.62	-10.7%	.001	10.7 (1)
	obj by propWd_cns	0.62	-10.8%	.000	12.6 (1)
	med by propSyl_ph	0.47	-16.0%	.000	18.1 (1)
	Constant	0.12	-	.000	26.0 (1)

Table 6.15: Factors which significantly affected the likelihood of a kontrast in the full regression model including all words (i.e. S+Ph+Prom, see Appendix E for description of features). The odds ratio for each feature in the model (Exp(B)), as well its significance (Sig) using the Wald statistic is given, along with the percentage change (P diff) in the likelihood of a kontrast with the presence of that level of that variable (or a one unit increase for continuous variables).

Tables 6.15 and 6.16 show the factors which significantly affect the likelihood of a kontrast for all words, and nuclear accented, words respectively in the final regression models, using the same format as in the previous experiments (see Appendices F.6 and F.7 for a listing of all levels of all parameters in each model).

K on N	Feat	Exp(B)	P diff	Sig	Wald (df)
Increase	new	1.87	11.5%	.000	23.7 (1)
	propWd_cl	4.24	21.3%	.000	77.3 (1)
	head_cns by propSyl_ph	2.76	16.9%	.016	5.8 (1)
	adjunct by prom_wd	1.38	6.4%	.009	6.8 (1)
	RB by prom_wd	2.16	13.7%	.000	26.0 (1)
	JJ by prom_wd	4.29	21.4%	.000	82.4 (1)
	PR by prom_wd	1.90	11.8%	.000	14.3 (1)
	VB by prom_wd	2.85	17.3%	.000	47.1 (1)
	NN by prom_wd	3.13	18.4%	.000	76.4 (1)
	DT by prom_wd	1.90	11.8%	.000	15.8 (1)
	numSyl_wd	1.45	7.3%	.000	42.9 (1)
Decrease	old	0.61	-11.5%	.000	13.8 (1)
	head_cns	0.32	-27.1%	.003	8.8 (1)
	adjunct by propWd_cns	0.47	-18.1%	.002	10.0 (1)
	obj by propWd_cns	0.62	-11.0%	.042	4.1 (1)
	Constant	0.08	-	.000	76.3 (1)

Table 6.16: Factors which significantly affected the likelihood of a kontrast on a nuclear accent in the full regression model (i.e. S+Ph+Prom, see Appendix E for description of features). The odds ratio for each feature in the model (Exp(B)), as well its significance (Sig) using the Wald statistic is given, along with the percentage change (P diff) in the likelihood of a kontrast with the presence of that level of that variable (or a one unit increase for continuous variables).

Looking at the all-word model first, we can see the types of semantic factors which make a kontrast more likely are generally as would be expected. A kontrast is substantially more likely if the word is a noun (NN) or an adjective (JJ). There is also, curiously, a more moderate increase if the word is a verb (VB) or a determiner (DT). The latter could be because these determiners are either included in kontrastive NPs, or are demonstrative. As expected, the likelihood of a kontrast increases as the proportion of the syntactic clause increases (*propWd_cl*), and decreases if the word is the head of a constituent (head_cns). Interestingly, it also increases if the word is mediated (med). This may be because common sub-types of mediated status, such as *set* or *situation*, are often kontrastive because they pick out parts of a general theme the speaker is talking about. It may also be because another

common subtype, *general*, picks out generally known entities that nevertheless behave in a similar way to *new* entities. The interaction between info status and info type was not significant, however. New entities were only significantly more likely to be kontrastive towards the end of a phrase (new by *propSyl_ph*). This could be because new elements were generally longer, and so only the end of the unit was stressed and therefore marked as kontrastive.

We can also see the interaction between semantic/syntactic properties and structural prominence. Generally a pronoun is not likely to be kontrastive. However, if it is accented this likelihood increases drastically (accq by PR and nuc by PR). While overall a constituent head is less likely to be kontrastive, this increases considerably as the proportion of the phrase increases (head_cns by *propSyl_ph*). That is, the effect of information structure within constituents seems to work differently than between constituents at the clause level; where, as we suggested before, syntax structure may be manipulated to put important information in nuclear position. Structural prominence also interacts with acoustic prominence: the likelihood of an adjunct being kontrastive increases with its prominence (adjunct by *prom_wd*). Further, nouns, which are already expected to be accented, must be in nuclear position *and* have increased prominence to increase the likelihood of a kontrast (nuc by NN by *prom_wd*).

Turning to the prediction model on nuclear accents, we know that if an element is nuclear, it is already highly likely to be kontrastive, so we are interested in what increases this likelihood. Again, we can see the effect of the overall likelihood of different elements being kontrastive: new elements (new) are more likely, old elements (old) less likely. Once more, likelihood increases with the proportion of the clause (*propWd_cl*) and decreases if the element is a head (head_cns). Most interesting is the interaction with part-of-speech type. We can see that increasing prominence increases likelihood for all types, but this is mediated by the likelihood that these elements would be kontrastive anyway. For instance, a pronoun is unlikely to be kontrastive, so it merely being nuclear makes a kontrast more likely. Therefore increasing acoustic prominence doesn't increase this probability as much (PR by *prom_wd*). On the other hand, an adjective is likely to be kontrastive, therefore the increase in acoustic prominence carries more information and increases this probability more (JJ by *prom_wd*).

Overall, these results uphold our hypothesis than kontrast is signalled by a combination of the inherent likelihood of different semantic/syntactic and prosodic structures being kontrastive, and by the acoustic prominence of individual words given their features in this structure.

6.3.2 K2: Kontrastive Accents Raised

In the second set of experiments, we look more closely at which factors significantly affect the major acoustic correlates of accents. We firstly look at whether there are significant

differences by accent status (pre-, post-, nuclear), and secondly, whether this is manipulated to show kontrast status.

6.3.2.1 Aim and Method

Our general claim and the specific hypotheses being tested in this experiment are:

- **General Claim:** On average, nuclear accents are more acoustically prominent than plain accents. However, this is manipulated to make kontrastive accents more acoustically prominent given their accent status. This may be to distinguish non-kontrastive accents, or to give a *restricted* kontrast reading.
- **Hypothesis 1:** Pre-, post- and nuclear accents have distinct acoustic profiles.
- **Hypothesis 2:** Kontrastive accents are more acoustically prominent than non-kontrastive accents by accent status.

Since, in this experiment, there were multiple, continuous dependent variables, we used several MANCOVAs to test the significance of the effects we were interested in. In relation to the first hypothesis our model included a single main factor of Accent Status (prenuclear, nuclear, postnuclear). For the second hypothesis, we tested the main factor Kontrast Status (konnp, konword, background) on each type of accent separately. In all models the following covariates were also included to control for the effect of accent position and overall phrase pitch and intensity on the dependent acoustic variables being tested: *propSyl_ph*, *numWd_ph*, *accsPh_inc*, *npmean_ph*, *nimean_ph* (see descriptions in Appendix E).

The acoustic correlates of accenting being tested were the same as those tested as correlates of prominence in the earlier experiments (e.g. see word acoustic features in Appendix E.2.6). However, only those variables which showed the greatest effects were included in the final accent status model, i.e. *npmax_wd*, *npqrange_wd*, *nimean_wd* and *dur_relSyl* (maximum pitch, inter-quantile pitch range, mean intensity and relative duration of the word). In the kontrast models, the effect on some of these variables was not significant, and so they were excluded. We also wanted to test the effect of accent status and kontrast on the position of the start of the rise to the accent peak (*naccL_time*) and the time of the peak (*naccH_time*), given the literature in Chapter 3 suggesting that these factors might affect accent shape. However, since the normalisation procedures for these measures were much more rigorous than for the other acoustic measures (see section 5.5), there were many fewer data points. Therefore, these variables were tested separately so as not to reduce the power of the model with the other variables. The data set was the same as in the previous experiments, excluding unaccented words. In the accent status models, there were 4992 words in

the general model and 3756 in the accent shape model. In the kontrast model on pre-nuclear accents, there were 1927 words in the general model, and 1437 on the shape model. In the kontrast model on nuclear accents, there were 2827 words in the general model and 1994 in the shape model. It was decided not to investigate post-nuclear accents since there were so few (243), and they seem to behave much less consistently than pre- or nuclear accents.

In the results below, the effects on pitch, intensity and duration measures are reported in normalised units, and it can be difficult to judge from these findings how perceptible these differences are. As a rough guide, recall that a one unit increase in normalised pitch represents a 10% increase in that speaker's logged pitch range (excluding outliers). Over all the speakers in the part of the corpus used for this study, average pitch range was 210 Hz for women and 95 Hz for men. Therefore a one unit increase is roughly 21 Hz for women and 9.5 Hz for men (though this is only approximate since pitch values were not evenly distributed over the range). Intensity values are relative to the mean intensity for all words by that speaker in that conversation, multipled by 10. So values above 10 are higher than average intensity. For the duration measure, one unit is 10ms, and this is divided by the number of syllables in the word, so the measure is approximately the change in duration per syllable.

6.3.2.2 Results and Discussion

Our first experiment looked at the effect of accent status on major acoustic correlates of prominence. Using a one factor multivariate ANCOVA, there was a highly significant main effect of Accent Status, as well as highly significant main effects of our five covariates which controlled for phrasal position and prominence ($p < 0.0001$, see Appendix F.8 for full multivariate significance test results). As can be seen in Table 6.17, the effect of Accent Status on each of the dependent variables was highly significant using *post-hoc* univariate tests (Bonferroni correction). We were particularly interested in whether pre-, post- and nuclear accents have distinct acoustic profiles; after controlling for known phrase level effects such as declination and final lowering, as well as the overall prominence of the phrase. Therefore, Table 6.17 also shows the estimated marginal means for each of our dependent variables, after factoring out the effect of our covariates. We also tested the effect of Accent Status on accent shape. Overall there was a slight effect on peak position (*naccH_time*) between nuclear and pre-nuclear accents in the direction expected, i.e. pre-nuclear peaks are later than nuclear peaks. However, there was no difference between nuclear and post-nuclear accents, and the effect of Accent Status on peak position in the post-hoc test was not at all significant ($p < 0.414$), therefore this result is not reported. There was also no consistent effect of Accent Status on the time of the minimum (*naccL_time*).

Multivariate Test of Significance					
				Pillai's, Hotelling's, Wilk's	$p < 0.0001$
Estimated Marginal Means and Univariate Tests					
Variable	**acc_stat**	**Mean**	**Std Err**	**F(2,4984)**	**Sig**
npmax_wd	**pre**	6.86	.045	62.4	.000
	nuc	7.07	.033		
	post	5.96	.100		
npqrange_wd	**pre**	1.70	.045	23.1	.000
	nuc	1.86	.034		
	post	1.18	.101		
nimean_wd	**pre**	10.07	.020	9.9	.000
	nuc	10.19	.015		
	post	10.15	.044		
dur_relSyl	**pre**	2.39	.035	39.4	.000
	nuc	2.80	.026		
	post	2.45	.079		

Table 6.17: Results from a MANCOVA showing the effect of accent status (*pre*-nuclear, *nuc*lear, *post*-nuclear) on the acoustic features of accented words (4992), including normalised maximum pitch (npmax_wd), quantile pitch range (npqrange_wd), mean intensity (nimean_wd) and relative duration (dur_relSyl). The estimated marginal means for each dependent by accent status are given (controlling for the proportion of syllables in the phrase so far, the total number of words in the phrase, the number of accents in the phrase so far, and the normalised mean pitch and intensity of the phrase). The standard error of each mean and the significance of the effect of accent status on each dependent using univarate tests, along with F-scores, are also reported.

The effect of accent status on each acoustic variable can be seen more clearly in Figure 6.2. This shows that accent status has a distinct effect on each one of these variables. Firstly, nuclear accents seem to be distinguished from other accents most clearly by their intensity and duration features. This is not an effect that is often cited in the literature (but see Kochanski et al. 2005); although, if it proves as robust as it seems to here, it could provide an answer to sceptics of the existence of nuclear accents, whose major evidence is the lack of distinct pitch contour effects on nuclear accents. We will return to this in the discussion at the end of the chapter. Interestingly, pre-nuclear accents are not louder and longer than

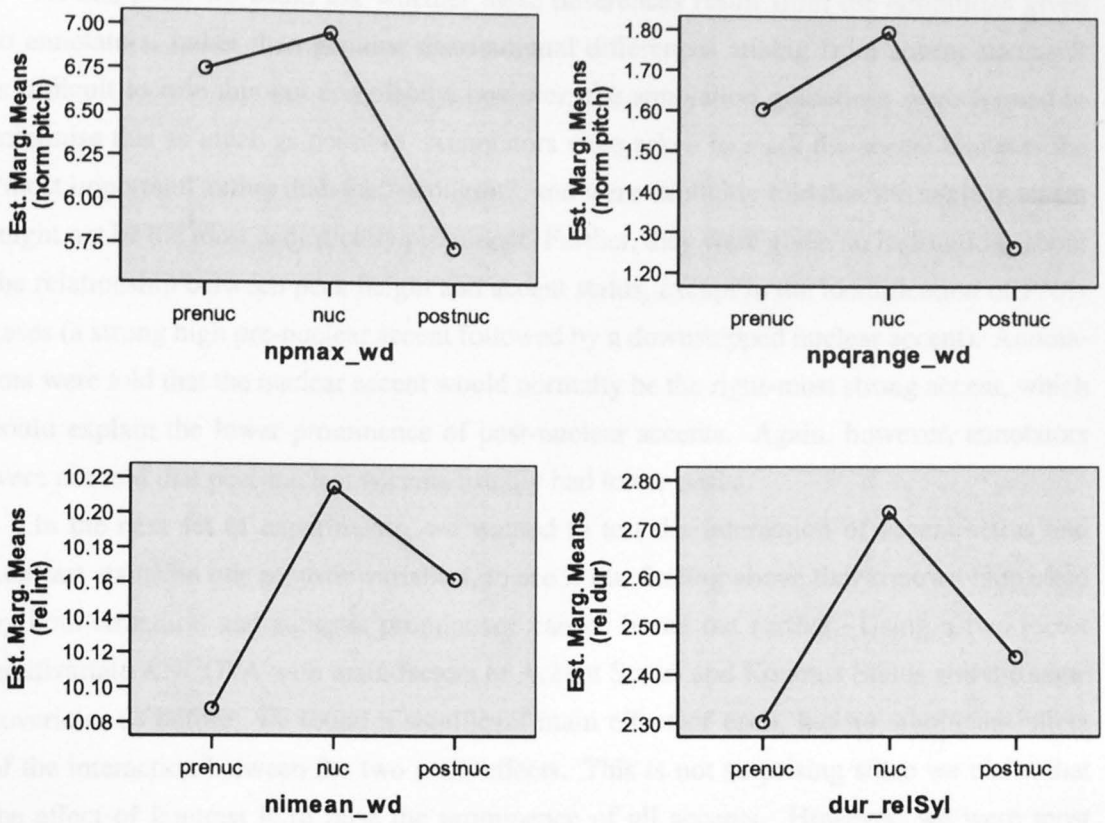


Figure 6.2: Graphs showing the effect of accent status (*prenuc*lear, *nuc*lear, *postnuc*lear) on the word-level acoustic features of accents (controlling for *propSyl_ph*, *numWd_ph*, *accPh_inc*, *npmean_ph* and *nimean_ph*), including normalised maximum pitch (npmax_wd), quantile pitch range (npqrange_wd), mean intensity (nimean_wd) and relative duration (dur_relSyl). Note the y-axis for each dependent shows normalised units for that feature (see text for interpretation).

post-nuclear accents, as might be expected, since they are generally heard as being more prominent. However, this backs up the claim made in section 3.1.2 that post-nuclear accents are primarily marked by increased duration and intensity, since the pitch range is reduced in the post-nuclear region. We can see that there is only a small difference between the peak heights of pre-nuclear and nuclear accents, while post-nuclear peaks are much lower. Finally, the pitch range of nuclear accents is, on average, larger than that of pre-nuclear accents. This is also consistent with our suggestion in section 3.2.4 that accent shape differences (which can most clearly be seen on large pitch movements) are primarily marked on nuclear accents.

We will return to this in our discussion of the next set of models.

At this point we could ask whether these differences result from the definitions given to annotators, rather than genuine distributional differences arising from accent status. It is difficult to rule this out completely, however, the annotation guidelines were framed to minimise this as much as possible. Annotators were asked to mark the accent that was the "most important" rather than the "strongest", and were explicitly told that the nuclear accent might not be the most acoustically prominent. Further, they were given no instructions about the relationship between peak height and accent status, except in the identification of PN/N cases (a strong high pre-nuclear accent followed by a downstepped nuclear accent). Annotators were told that the nuclear accent would normally be the right-most strong accent, which could explain the lower prominence of post-nuclear accents. Again, however, annotators were not told that post-nuclear accents usually had lower peaks.

In the next set of experiments, we wanted to test the interaction of accent status and kontrast status on our acoustic variables, to see if our finding above that kontrast is marked by both structural and acoustic prominence can be borne out further. Using a two factor multivariate ANCOVA with main factors of Accent Status and Kontrast Status and the same covariates as before; we found a significant main effect of each, but no significant effect of the interaction between the two main effects. This is not surprising since we claim that the effect of kontrast is to raise the prominence of all accents. However, we were most interested to see if the marking of kontrast status concurred with our claims above about how prominence is marked on each type of accent. Therefore we do not report this finding and move on to a series of MANCOVAs showing the effect of kontrast status on pre-nuclear and nuclear accents separately.

Table 6.18 shows the results of a one factor multivariate ANCOVA on pre-nuclear accents with Kontrast Status as the main effect and the same covariates as before. All factors were highly significant ($p < 0.0001$, see Appendix F.9 for full test results). Post-hoc tests showed there was only a significant effect of Kontrast Status on *npmax_wd* and *dur_relSyl*, so the other dependent variables were excluded. Again, estimated marginal means and univariate test results are reported. A separate MANCOVA was done to test the effect of Kontrast Status on *naccL_time* and *naccH_time*. However, there were no significant or suggestive differences.

Figure 6.3 shows the effect of kontrast status on pre-nuclear accents, after controlling for phrase position and prominence. These effects work in the direction expected, with *konword* elements being higher and longer than *background* elements. *konnp* elements are approximately mid-way being these: probably because annotated NPs contain words which are marked as the head of the kontrast phrase, and other words which are not prosodically prominent, but still fall within the scope of the head. If we compare the results for *npmax_wd*

Multivariate Test of Significance					
				Pillai's, Hotelling's, Wilk's	$p < 0.0001$
Estimated Marginal Means and Univariate Tests					
Variable	**kon_stat**	**Mean**	**Std Err**	**F(2,1919)**	**Sig**
npmax_wd	**kword**	7.44	.061	5.4	.005
	knp	7.27	.077		
	bkgd	7.19	.043		
dur_relSyl	**kword**	2.30	.039	24.1	.000
	knp	2.24	.049		
	bkgd	2.00	.027		

Table 6.18: Results from a MANCOVA showing the effect of kontrast status (*k*ontrastive *word*, *k*ontrastive *np*, background (*bkgd*)) on the acoustic features of words with pre-nuclear accents (1927), including normalised maximum pitch (npmax_wd) and relative duration (dur_relSyl). The estimated marginal means for each dependent by kontrast status are given (controlling for the proportion of syllables in the phrase so far, the total number of words in the phrase, the number of accents in the phrase so far, and the normalised mean pitch and intensity of the phrase). The standard error of each mean and the significance of the effect of kontrast status on each dependent using univarate tests, along with F-scores, are also reported.

in Figure 6.2 and 6.3, we can see that the estimated mean for pre-nuclear konwords in substantially higher (7.44) than the overall mean for nuclear accents (7.07), consistent with our prediction that kontrast can raise the prominence of pre-nuclear accents without changing the perception of the pre-nuclear/nuclear distinction. Overall, nuclear accents are still longer, however.

Table 6.19 shows the results of a one factor multivariate ANCOVA on nuclear accents with Kontrast Status as the main effect and the same covariates. All factors were significant ($p < 0.05$, see Appendix F.10 for full result listing). Post-hoc tests showed there was only a significant effect of Kontrast Status on *npmax_wd*, *npqrange_wd* and *dur_relSyl*, so *nimean_wd* was excluded. Again, estimated marginal means and univariate tests are reported. A separate MANCOVA was done which found a significant main effect of Kontrast Status on *naccH_time* and *naccH*, but not *naccL_time*. All factors were highly significant ($p < 0.01$, see Appendix F.11 for full result listing). Table 6.19 also shows estimated marginal means and *post-hoc* univariate tests for this model.

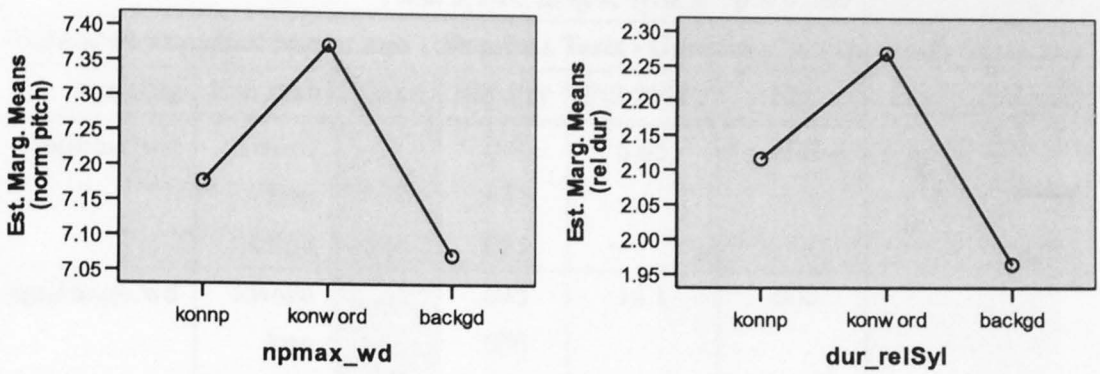


Figure 6.3: Graphs showing the effect of kontrast status (*k*ontrastive *word*, *k*ontrastive *np*, background (*bkgd*)) on the word-level acoustic features of *pre-nuclear* accents (controlling for *propSyl_ph*, *numWd_ph*, *accPh_inc*, *npmean_ph* and *nimean_ph*), including normalised maximum pitch (npmax_wd) and relative duration (dur_relSyl). Note the y-axis for each dependent shows normalised units for that feature (see text for interpretation).

Figure 6.4 shows the general acoustic features and peak features of nuclear accents which showed a significant effect by kontrast status. Once more, we can see that *npmax_wd* is higher for *konword*, and in this case also *konnp*, than *background*. The estimated marginal mean for maximum pitch is lower than the corresponding value for pre-nuclear accents. This result follows on from our general story: since nuclear accented material is already likely to be kontrastive, speakers do not need to use acoustic prominence to mark kontrast. On the other hand, since pre-nuclear accents do not usually directly convey meaning, extra acoustic prominence needs to be used to show kontrast status.

Following on from our finding that *npqrange* and *dur_relSyl* are significantly greater for nuclear accents, we find that they also vary significantly with kontrast status. *konword* elements have substantially greater pitch range and duration than the average marginal means for nuclear accents (cf. Table 6.17). This concurs with the idea that the way prosodic prominence is marked varies according to place in the prosodic structure. In addition to this, we can see that there is a significant effect on peak alignment, with the peak of both *konnp* and *konword* nuclear accents being later than on *background* accents. In section 3.2.4 we suggested that one function of 'emphatic accents' is to force a *restricted* kontrast interpretation for the accented element. We also presented evidence that late peaks may be a perceptual substitute for high peaks to mark an accent as particularly emphatic. Although it is not con-

Multivariate Test of Significance					
				Pillai's, Hotelling's, Wilk's	$p < 0.000$
Estimated Marginal Means and Univariate Tests - General					
Variable	**kon_stat**	**Mean**	**Std Err**	**F(2,2819)**	**Sig**
npmax_wd	**kword**	6.92	.041	8.65	.000
	knp	6.89	.041		
	bkgd	6.65	.051		
npqrange_wd	**kword**	2.11	.045	14.1	.000
	knp	1.85	.070		
	bkgd	1.74	.056		
dur_relSyl	**kword**	3.07	.036	7.97	.000
	knp	2.91	.056		
	bkgd	2.85	.045		
Estimated Marginal Means and Univariate Tests - Peak					
Variable	**kon_stat**	**Mean**	**Std Err**	**F(2,1987)**	**Sig**
naccH	**kword**	7.14	.048	11.7	.000
	knp	6.92	.077		
	bkgd	6.76	.063		
naccH_time	**kword**	6.17	.084	5.9	.003
	knp	6.25	.133		
	bkgd	5.75	.108		

Table 6.19: Results from a MANCOVA showing the effect of kontrast status (*k*ontrastive *word*, *k*ontrastive *np* , background (*bkgd*)) on the acoustic features of words with nuclear accents. Separate models are reported showing the effect on general features (2827 tokens), including normalised maximum pitch (npmax_wd), quantile pitch range (npqrange_wd) and relative duration (dur_relSyl); and on peak features (1994 tokens), including normalised peak height (naccH) and location (naccH_time). The estimated marginal means for each dependent by kontrast status are given (controlling for the proportion of syllables in the phrase so far, the total number of words in the phrase, the number of accents in the phrase so far, and the normalised mean pitch and intensity of the phrase). The standard error of each mean and the significance of the effect of kontrast status on each dependent using univarate tests, along with F-scores, are also reported.

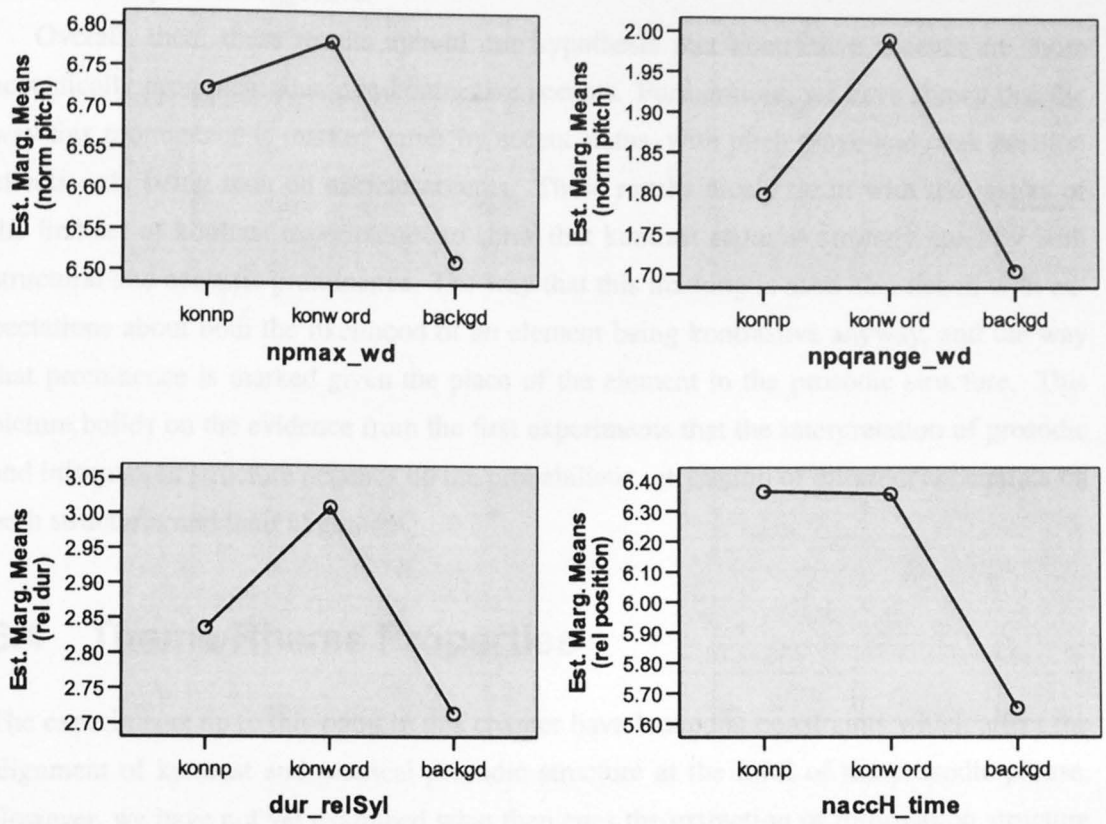


Figure 6.4: Graphs showing the effect of kontrast status (*k*ontrastive *word*, *k*ontrastive *np*, background (*bkgd*)) on the word-level acoustic features of *nuclear* accents (controlling for *propSyl_ph*, *numWd_ph*, *accPh_inc*, *npmean_ph* and *nimean_ph*), including normalised maximum pitch (npmax_wd), quantile pitch range (npqrange_wd), relative duration (dur_relSyl) and normalised peak location (naccH_time). Note the y-axis for each dependent shows normalised units for that feature (see text for interpretation), and that the peak location estimates come from a different model to the other three dependents.

clusive, these findings are suggestive of this relationship. We will return to this in the general discussion. It is also interesting to note that pitch range and peak position only seem to be a significant factor on nuclear accents. This is consistent with our suggestion in section 3.2.4 that nuclear accents are not only important structurally, but most of the accent shape variation that annotation systems like ToBI try to capture is only important on nuclear accents, i.e. variation that changes the illocutionary and affective connotations of the phrase. Of course, these results are far from conclusive on this point, but since substantial local varia-

tion in pitch range is a pre-requisite to being able to express standardly accepted variations in accent shape, it is suggestive.

Overall, then, these results uphold our hypothesis that kontrastive accents are more acoustically prominent than non-kontrastive accents. Furthermore, we have shown that the way this prominence is marked varies by accent status, with pitch range and peak position effects only being seen on nuclear accents. These results nicely tie in with the results of the first set of kontrast experiments, to show that kontrast status is strongly cued by both structural and acoustic prominence. The way that this marking is used also ties in with expectations about both the likelihood of an element being kontrastive anyway, and the way that prominence is marked given the place of the element in the prosodic structure. This picture builds on the evidence from the first experiments that the interpretation of prosodic and information structure depends on the probabilistic interaction of different constraints on both structures and their alignment.

6.4 Theme/Rheme Properties

The experiments up to this point in this chapter have looked at constraints which affect the alignment of kontrast and metrical prosodic structure at the level of the prosodic phrase. However, we have not yet examined what then cues the extraction of information structure across prosodic phrases. In section 3.1.2, we laid out evidence that prosodic phrasing is recursive, and so the perception of nuclear prominence can carry across phrases. We claimed that the acoustic signalling of this structure mirrors that at the phrase level, i.e. the last of roughly equally prominent nuclear accents will be perceived as the nucleus of the higher phrase. For a phrase to be perceived as subordinate in relation to the previous one, its nuclear accent must be substantially reduced.

In Chapter 4 we showed how this links in with the perception of information structure, i.e. thematic kontrasts are less prosodically prominent than rhematic kontrasts, using controlled phonetic experiments. The aim of the final experiment here is to test whether the results there hold up for the much more varied real speech data in our corpus. As we discussed in the last chapter, we decided it would be too difficult for annotators to consistently mark phrasing structure, and hence nuclear accenting, at a level higher than the basic prosodic phrase. Therefore we could not test our claim directly in terms of nuclear accent and boundary placement. However, we could look at the relative prominence of nuclear accents in theme and rheme phrases within the same information unit, to see if the expected relationship holds. Further, as was explained in the last chapter, only one Switchboard conversation has currently been annotated for theme/rheme status, so the data set is small; especially considering it

is uncontrolled spontaneous speech. We have therefore included some results which are indicative but not statistically significant.

6.4.1 Aim and Method

Our general claim and the specific hypotheses being tested in this experiment are:

- **General Claim:** Themes are less prosodically prominent than rhemes. When themes and rhemes appear in separate prosodic phrases, their status will be indicated by a combination of structural and acoustic prominence. That is, in theme/rheme order, status is shown by the rheme being the last of equally acoustically prominent phrases. In rheme/theme order, however, thematic status is shown by the lower acoustic prominence of the phrase in relation to the rheme.

- **Hypothesis 1:** The heads of thematic phrases (i.e. nuclear accents) will be less acoustically prominent than the heads of rhematic phrases only when the theme phrase occurs after the rheme phrase.

- **Hypothesis 2:** The difference between the acoustic features of the heads of rheme and theme phrases is positive only in rheme/theme order (in the same information unit). In this case, rheme heads are more prominent.

As in the last experiment, we were interested in effects on the acoustic correlates of prominence, so we used a series of ANOVAs to test the significance of different effects. In the first set of experiments, we wanted to see if there was an absolute difference between the acoustic profiles of the nuclear accents of theme and rheme phrases. Therefore we tested main effects of Theme/Rheme Status and Place. Place recorded whether the theme/rheme phrase came before or after its information unit pair. In the case of discontinuous phrases, e.g. theme-rheme-theme, the first phrase was classed as being *first*, and the last two as being *second*. Of the covariates in the last experiments, only *propSyl_ph* proved to be significant here, probably because of the small sample size, so the others were excluded. In the second set of experiments, we looked at the *difference* between acoustic features of paired theme and rheme phrases. Therefore the data set consisted of paired samples of acoustic variables. Here we were looking for a main effect of Order, i.e. theme/rheme versus rheme/theme. Discontinuous phrases were included twice, once for each possible linear pairing. None of the previous covariates were significant so they were excluded.

Our data set came from the one conversation in which prosodic phrases had been annotated as thematic or rhematic. As discussed in section 5.7, phrases which were disfluent or

Multivariate Test of Significance						
				Pillai's, Hotelling's, Wilk's	$p < 0.007$	
Estimated Marginal Means and Between-Subjects Effects						
Variable	**tr_place**	**tr_stat**	**Mean**	**Std Err**	**F(3,256)**	**Sig**
nimean_wd	**first**	**theme**	10.25	.095	3.0	.029
		rheme	9.78	.209		
	second	**theme**	9.73	.226		
		rheme	10.27	.087		
dur_relSyl	**first**	**theme**	2.12	.108	2.4	.068
		rheme	2.17	.238		
	second	**theme**	2.26	.258		
		rheme	2.50	.099		

Table 6.20: Results from a MANCOVA showing the effect of Place (first/second) and Status (theme/rheme) on the acoustic features of accented words in the subset annotated for theme/rheme (261 tokens), including normalised intensity (nimean_wd) and relative duration (dur_relSyl). The estimated marginal means for each dependent by Place and Status are given (controlling for the proportion of syllables in the phrase so far). The standard error of each mean and the significance of the effect of Place and Status on each dependent using univarate tests, along with F-scores, are also reported.

which contained only discourse markers, e.g. *anyway*, without any clear information structure, were excluded. Of the remainder, only some were marked as 'paired', i.e. separate theme and rheme phrases that clearly formed part of the same information unit. Since we were most interested in the main effects of Place and Order, the rest of the phrases were excluded. This left a total of 109 nuclear-accented words in the first experiment. We decided to also test all accented words to increase the power of the test, a total of 261 words. As before, we tested the effect on accent shape (*naccL_time* and *naccH_time*) separately as there were fewer data points: 88 nuclear accents. In the second experiment, there were 39 theme-rheme pairs with non-missing values for the relevant acoustic variables; and 27 in the accent shape comparison.[2]

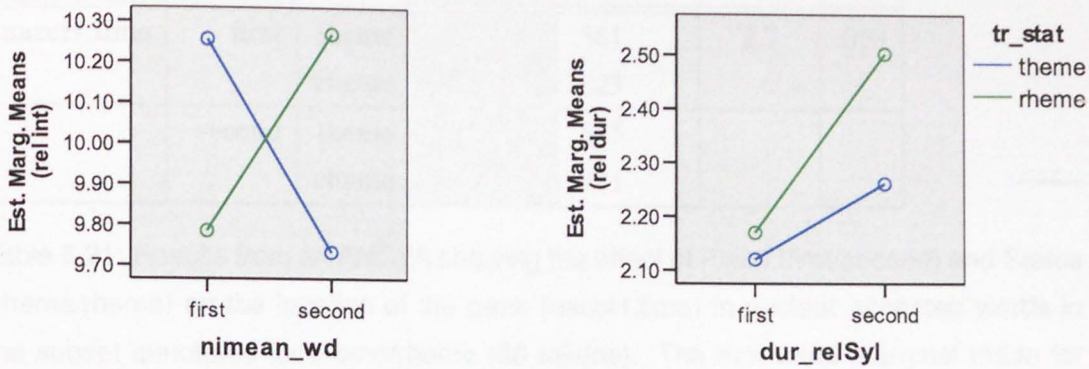


Figure 6.5: Graphs showing the effect of Status (theme/rheme) and Place (first/second) on the word-level acoustic features of accents (controlling for *propSyl_ph*), including normalised mean intensity (nimean_wd) and relative duration (dur_relSyl). Note the y-axis for each dependent shows normalised units for that feature (see text for interpretation).

6.4.2 Results and Discussion

Initially we tested a two factor multivariate ANCOVA with main effects of Status and Place on the acoustic features of nuclear accents. The main effects of Status and Place did not even approach significance, but their interaction neared significance. Therefore we ran all further tests with a single main factor of TR_Status*Place. There was no significant effect of this interaction on nuclear accents. However, there was a significant main effect when all accents were included ($p < 0.007$, $F(6,512) = 0.07$). There was also a significant main effect of *propSyl_ph* ($p < 0.0001$, $F(2,255) = 0.12$). As can be seen in Table 6.20, between-subjects effects tests showed that, of the acoustic variables tested, there was only a significant effect on *nimean_wd* and a marginally significant effect on *dur_relSyl*. This can be seen graphically in Figure 6.5. When the theme phrase comes before the rheme phrase, accents in theme phrases are more intense than accents in rheme phrases. This is reversed when the theme follows the rheme. With duration, on other hand, rhematic accents are always longer than thematic accents, but this effect is much greater when the theme follows the rheme. Using a univariate ANOVA, we found a nearly significant effect of TR_Status*Place on peak position (*naccH_time*), see Table 6.21. There was no effect on *naccL_time*, nor of any of the earlier

[2]There were also fewer values in this comparison because only the last nuclear accent in each theme and rheme phrase was compared; whereas in the first experiment all nuclear accents in contiguous theme or rheme phrases were included.

Estimated Marginal Means and Between-Subjects Effects						
Variable	**tr_place**	**tr_stat**	**Mean**	**Std Err**	**F(3,84)**	**Sig**
naccH_time	**first**	**theme**	5.66	.361	2.7	.053
		rheme	5.28	.823		
	second	**theme**	7.12	.555		
		rheme	6.59	.271		

Table 6.21: Results from an ANOVA showing the effect of Place (first/second) and Status (theme/rheme) on the location of the peak (naccH_time) in nuclear accented words in the subset annotated for theme/rheme (88 tokens). The estimated marginal mean for the dependent by Place and Status is given. The standard error of each mean and the significance of the effect of Place and Status on the dependent, along with F-scores, are also reported.

tr_order	**Mean**	**Std Err**	**t (df)**	**Sig**
theme-rheme	2.39	.138	3.12 (49)	.003
rheme-theme	3.02	.172		

Table 6.22: The mean and standard error (std err) of the *difference* between the relative duration (dur_relSyl) of rheme and theme nuclear accented words (39 tokens). The significance of the effect of theme/rheme order is shown using a t-test, the t statistic and degrees of freedom (df) are also reported.

covariates. As we can see in Figure 6.6, in both orders peaks on thematic nuclear accents are later than peaks on rhematic nuclear accents. The difference is slightly greater when the theme follows the rheme. This would suggest some kind of subtle shape variation between theme and rheme accents, we will return to this below. These findings are in the direction that we would expect given our first hypothesis. We can only suspect that the lack of significant effects on pitch features is due to the lack of data, particularly since the model could not account for global changes in pitch levels between phrases.

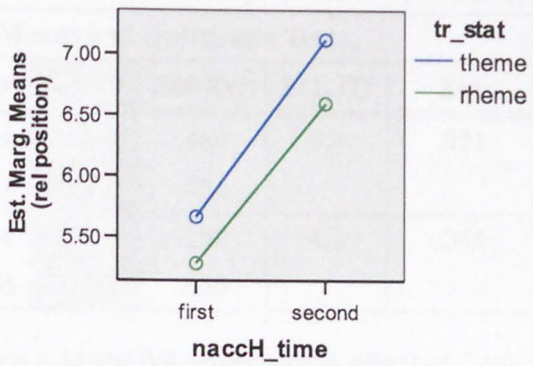


Figure 6.6: Graph showing the effect of Status (theme/rheme) and Place (first/second) on the location of the peak (naccH_time) in nuclear accents. Note the y-axis shows normalised units (see text for interpretation).

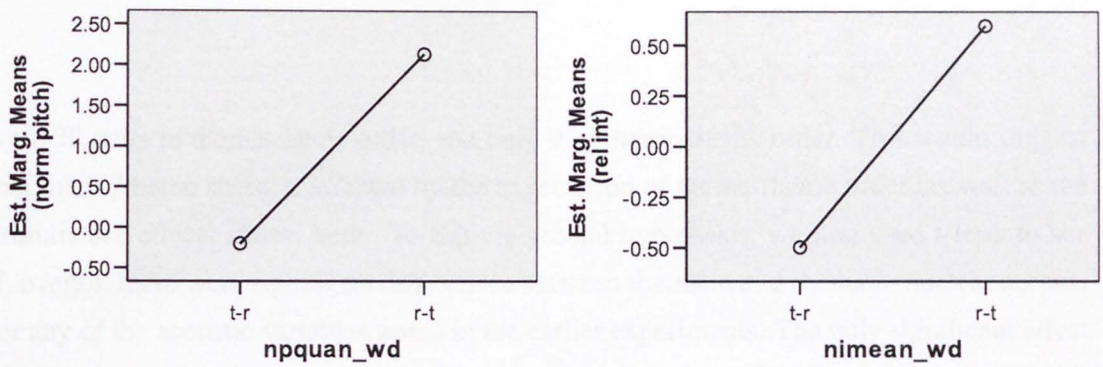


Figure 6.7: Graphs showing the effect of Order (*t*heme-*r*heme versus *r*heme-*t*heme) on the *difference* between the acoustic features of rheme and theme nuclear accented words, including normalised quantile pitch (npquan_wd) and mean intensity (nimean_wd). Note the y-axis for each dependent shows the difference between normalised units for that feature (see text for interpretation).

In relation to the second hypothesis, we looked specifically at the *differences* between the acoustic features of paired themes and rhemes; since our claim is not that themes are less prominent than rhemes overall, but that they are relatively less prominent than their paired rheme in the same information unit. The data set was very small, but in itself revealing. There

Multivariate Test of Significance					
			Pillai's, Hotelling's, Wilk's		$p < 0.015$
Estimated Marginal Means and Univariate Tests					
Variable	**tr_order**	**Mean**	**Std Err**	**F(1,37)**	**Sig**
npquan_wd	**t-r**	-0.20	.466	5.8	.021
	r-t	2.14	.851		
nimean_wd	**t-r**	-0.50	.252	4.3	.044
	r-t	0.60	.460		

Table 6.23: Results from a MANOVA showing the effect of Order (*t*heme-*r*heme versus *r*heme-*t*heme) on the *difference* between the acoustic features of rheme and theme nuclear accented words in the subset annotated for theme/rheme (39 tokens), including normalised quantile pitch (npquan_wd) and mean intensity (nimean_wd). The estimated marginal mean for each dependent by Order is given. The standard error of each mean and the significance of the effect of Order on each dependent, along with F-scores, are also reported.

were 30 pairs in theme-rheme order, and only 9 in rheme-theme order. This would suggest that theme/rheme status is affected by the expectation of theme-rheme order, as well as the prominence effects shown here. To test the second hypothesis, we first used t-tests to see if, overall, there were significant differences between thematic and rhematic nuclear accents for any of the acoustic variables tested in the earlier experiments. The only significant effect was for *dur_relSyl*, rhemes are longer than themes (see Table 6.22). This nicely ties in with the finding in the last experiment that the most distinctive feature of nuclear accents is their duration. It follows that this effect would hold between phrases as well. Generally, then, there were no overall effects of theme/rheme status on prominence, but we needed to see if there would be effects in rheme-theme order. Using a one factor multivariate ANOVA, we found a main effect of Order ($p < 0.015$, $F(2,36) = 0.26$). As can be seen in Table 6.23, the effect was significant on *npquan_wd* and *nimean_wd*. Using a separate ANOVA, the effects on accent shape were not significant. However, there were only 27 data points and the effect was in the expected direction, so we include these results as well (see Table 6.24).

We can see the effects on acoustic features graphically in Figure 6.7. For each variable, a positive value shows the rheme was higher than the theme. Here we do see an effect on pitch: *npquan_wd* in theme accents is slightly greater than for rheme accents in theme-rheme order, but rhemes are substantially higher than themes when the order is reversed. Our finding

Multivariate Test of Significance					
			Pillai's, Hotelling's, Wilk's		$p < 0.380$
Estimated Marginal Means and Univariate Tests					
Variable	**tr_order**	**Mean**	**Std Err**	**F(1,25)**	**Sig**
naccL_time	**t-r**	0.28	1.64	0.32	.573
	r-t	-1.56	2.78		
naccH_time	**t-r**	0.49	.598	2.02	.167
	r-t	-1.18	1.01		

Table 6.24: Results from a MANOVA showing the effect of Order (*t*heme-*r*heme versus *r*heme-*t*heme) on the *difference* between the peak features of rheme and theme nuclear accented words in the subset annotated for theme/rheme (27 tokens), including normalised L and H locations (naccL_time and naccH_time). The estimated marginal mean for each dependent by Order is given. The standard error of each mean and the significance of the effect of Order on each dependent, along with F-scores, are also reported.

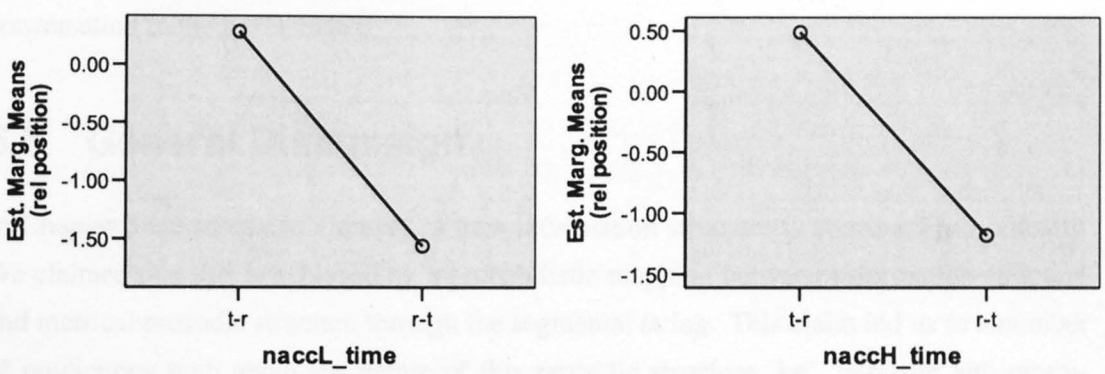


Figure 6.8: Graphs showing the effect of Order (*t*heme-*r*heme versus *r*heme-*t*heme) on the *difference* between the peak features of rheme and theme nuclear accented words, including normalised low (naccL_time) and peak (naccH_time) location. Note the y-axis for each dependent shows the difference between normalised units for that feature (see text for interpretation).

above that relative intensity is reversed in theme-rheme and rheme-theme order is confirmed. However, the effect on duration was not significant. Finally, we can see in Figure 6.8 that L and H tend to be later in theme accents than rheme accents only when themes follow rhemes.

However, this result is not significant and so needs to be treated with much caution.

Overall, these results consolidate the findings in Chapter 4 that thematic nuclear accents are less prosodically prominent than rhematic nuclear accents. This is shown by ordering in theme-rheme order, and by the theme accent being much less acoustically prominent in rheme-theme order. We found an overall effect of Status*Place on intensity and duration; and crucially, an effect on pitch and intensity difference in our paired samples, confirming our hypotheses. That these effects held is particularly note-worthy given the small and uncontrolled nature of the data set.

Like in the first production experiment in Chapter 4, we found a small but consistent effect of accent shape, particularly H alignment in this case: the peak of theme accents is later than for rheme accents. We saw then that this effect did not seem to hold up in the accompanying perception experiment. But these results do seem to suggest that there is a real effect here that needs to be explained. It could be, as we have suggested before, that thematic kontrasts are more likely to have a *restricted* kontrast reading, and therefore the later peak serves to emphasise this. Or it could arise from other affective connotations related to thematicity that we are not controlling for. Or this could in fact be a subtly distinct 'theme accent'. We will go into this question further looking at specific examples from this conversation in the next chapter.

6.5 General Discussion

In Chapter 3 we advanced a theory of how information structure is conveyed prosodically. We claimed that this is achieved by a probabilistic mapping between information structure and metrical prosodic structure through the segmental string. This claim led us to a number of predictions both about the nature of this prosodic structure, i.e. phrasing and prominence, and its relationship with the basic units of information structure, namely kontrast and theme/rheme status within syntactic structure. Using our annotated corpus, we have been able to test some of these predictions, and the results have been consistent with our claims. We firstly looked at features which are useful for predicting phrase breaks. We found, as expected, that clause and constituent type and structure act as strong constraints on phrasing. But this is mediated by positional and rhythmical constraints on the prosodic structure itself; and by kontrast and information status (although this was not as clearly captured by our features). Along with the finding that accentual features did not substantially improve phrase break prediction; these findings formed the foundation for our next series of studies, which presumed that prosodic phrases form the basic unit for the perception of prominence structure, and correlate strongly with information units. In the next series of studies we looked

at the features which were most useful for predicting levels of prominence, i.e. whether a word is unaccented, plain or nuclear accented. These studies showed a number of trends consistent with our claims: firstly, plain and nuclear accents can best be distinguished by their phrasal properties, showing the strong constraint of phrase structure, in particular the right-branching bias, on prominence perception. Secondly, nuclear accents can be reliably predicted by semantic/syntactic features, particularly kontrast, while other accents cannot. Plain accents may be more likely, however, if they occur on syntactically 'strong' words, e.g. nouns or objects, where these features were not significant for nuclear accents. This is consistent with our claim that there is a strong constraint aligning nuclear accents with kontrasts, while other accents may appear for rhythmical reasons, or because of low-level syntactic features, i.e. they are not usually directly involved in conveying information structure. Lastly, plain accent prediction substantially improved with the inclusion of word-level acoustic features, while nuclear accent prediction did not. This again shows that most plain accents are not directly 'meaningful', and therefore are not well predicted by other features.

The third set of experiments looked more closely at features which predict kontrast status. We found, as expected, that kontrast status was strongly cued by nuclear prominence; and that kontrast was likely to occur on semantically and syntactially 'strong' words. However, our results went further, allowing us to show how kontrast is conveyed by prominence given that the kontrast/nuclear accent relationship is not exact. We showed that a kontrast is more likely the more prominent a word is, given how prominent it is expected to be. That is, a kontrast is more likely if the word is in a structurally more prominent position than would be expected from its syntactic/information status properties; and if it is more acoustically prominent than would be expected given this and its structural prominence. Further, we showed that the acoustic correlates of this increased prominence vary depending on the expected acoustic correlates of different parts of the prosodic structure. Finally, our last, smaller, experiment confirmed the result in Chapter 4 that thematic nuclear accents are less acoustically prominent than rhematic nuclear accents in rheme-theme order but not in theme-rheme order. This is consistent with our claim that nuclear prominence can hold over multiple phrases, and that this is exploited to distinguish themes from rhemes by making themes less prosodically prominent than rhemes.

More generally, these results are encouraging for the overall prosodic framework set up in Chapter 3. We found clear, consistent acoustic differences between pre-, post- and nuclear accents (see Figure 6.2). In particular, nuclear accents are longer and have greater pitch range than other accents. In fact, our results could be seen as support for Kochanski et al.'s (2005) contention that loudness and duration are the primary acoustic correlates of prominence; and that $f0$ is much less important than previously assumed. Moreover, we saw earlier that these

acoustic features are not necessary for the prediction of nuclear accents, whose placement is highly constrained by phrase structure. This concurs nicely with our claim that nuclear accent perception is based on the expectation of the most prominent accent in a phrase: a nuclear accent does not actually have to be more acoustically prominent than pre-nuclear accents in the same phrase to be perceived as nuclear. Therefore acoustic prominence can be independently manipulated to convey detail about information structure (and possibly affective and illocutionary connotations of the phrase).

The clear divide in the reliability which which nuclear and plain accents can be predicted by semantic and syntactic features further strengthens evidence of a metrical structure. As we discussed in section 3.1.2, under the metrical view, pre-nuclear accents are a manifestation of strong nodes in the pre-nuclear region of the phrase. Depending on the length of the phrase, there may be several levels of relative prominence in this region, and so the phenomenon of 'accenting' is in fact a somewhat arbitrary cut-off point among these strong nodes. Hence the need for annotators to have a category of 'weak' accents. It follows from this it may be very difficult to recognise these 'accents' reliably from semantic/phrasal features, since their definition is somewhat arbitrarily related to the manner of their acoustic expression. Further, as we set out in Chapter 3, since English is a 'stress-timed' language, we expect to find not only the kontrast/nuclear accent constraint, but a more general constraint against making 'weak' elements stressed. This is nicely held up here, as we found plain accents were more likely if they fell on 'strong' syntactic elements. On the other hand, these results are very difficult to explain under the view that accents are independent events motivated by the semantic properties of the word; as some of the work reviewed in Chapters 2 and 3 claims. There would be no motivation for the semantic, syntactic, phrase structural and acoustic distinctions between plain and nuclear accents which we have found.

These results raise as many questions as they answer about the relationship between prosodic phrasing, syntactic phrasing and information structure. We have found that phrasing structure is strongly constrained by syntactic structure, with a less discernible effect of kontrast status. But we have also shown that nuclear accent placement is strongly constrained by both kontrast status and phrase structure. Taken together, these results would seem to indicate that it is syntax structure which is being manipulated in order to place kontrastive elements in prosodically strong positions. This would be an interesting suggestion since English is usually taken to be a language with relatively fixed syntax structure, so that prosody is varied to show information structure. This may be less true that has been thought. Certainly, this idea is suggestive of a system like Combinatory Categorial Grammar (Steedman 2001), where basic information units (theme/rheme) directly determine parsing and the nature of syntactic constituency. However, these questions are somewhat tangental

to our current purposes.

Overall, the success of these experiments can also be seen as confirmation of the reality of kontrast as a basic property of information structure. Kontrast *type* was only a minor factor in one of the semantic feature models, and not a significant feature in any of the others. This suggests that in marking instances of the various kontrast types, our annotators were identifying instances of a homogenous phenomenon. Unfortunately, these results are of limited use on the question of the reality of *restricted* kontrast, i.e. whether increased prosodic prominence makes the alternative set of the kontrast more salient; and the related question of the reality of 'emphatic accents', i.e. whether this increased prominence is gradient and/or categorical. We saw that increased acoustic prominence increases the likelihood of a kontrast, given the other features of the word. Further, we saw that peaks on kontrastive nuclear accents are consistently later than on backgrounded nuclear accents. Peak delay was one of the possible features of emphatic accents noted in section 3.2.4. However, the degree of peak delay found here, i.e. from an average of just before the middle of the stressed syllable to just after, is not consistent with the categorical effects noted in the literature, which involve at the minimum half-syllable differences. Therefore, these results seem more consistent with the view that a kontrast interpretation, and therefore presumably a *restricted* kontrast interpretation, is more likely the more prominent the accent; and one of the correlates of increased prominence on nuclear accents is gradient peak delay.

Information status features performed disappointingly poorly over all of the different models. We might have expected to find plain accents were less likely if the word was *old*, consistent with evidence of 'deaccenting givenness' presented in section 2.2.2. Further, there was no significant interaction between information status and the prominence of accents overall or by accent status, whereas some work in the literature predicts a reduction in prominence by increasing degrees of givenness (see section 2.2.2.2). This could be because information status was annotated using text only strictly on the basis of discourse-level givenness. As we discussed in Chapter 2, discourse givenness is related to, but can lead to quite different predictions from, relative Givenness, i.e. given in relation to the current proposition. The latter may be more relevant to prosodic prominence. It could also be that the semantic properties of the information status coding were already captured by the kontrast coding, and the effects were too hard to separate. In section 2.2.2.2, we also reviewed several studies which claimed prosodic prominence is related to the accessibility, predictability and/or informativity of a referent in a much more linear way than is suggested here. Our results here suggest that such reduction may form a separate 'stream' to the expression of prominence structure. For instance, we saw that Bard et al. (2000) claim reduction of intelligibility due to (discourse) givenness is due to automatic priming processes. Our finding

that information status was not a significant predictor of structural prominence suggests that such reduction does not directly affect the appearance of accents. Unfortunately, our fully annotated corpus is currently not big enough to build reliable language and semantic informativity models. This would enable us to directly compare the performance of our high-level and such low-level features in predicting prosodic events and overall prominence. However, there are currently plans to expand the data set, and such a comparison would certainly be worthwhile.

Outside of purely linguistic concerns, our results have implications for prosodic event prediction systems aimed at improving speech synthesis or language understanding systems. Our findings suggest that the most important variables to control in order to convey information structure are phrase boundary and nuclear accent placement. Other pitch accents may be inserted more freely based on low-level features such as part-of-speech and the number of words since the last accent. This is considerably different from many current systems which take phrase break and pitch accent prediction to be essentially unrelated processes; and which make no distinction between plain and nuclear accents. Kontrast has also been shown to be an important feature. We discuss how this work could be used to improve speech synthesis in section 8.3.

Finally, a note on the limitations of these experiments. Any model of corpus data is only a model of language output. Throughout this chapter, we have assumed that robust features of this output can be taken as evidence of language production and perception processes. However, it should be remembered that the link is not direct, and will need to be confirmed using complementary experimental methods such as phonetic production and perception experiments which are beyond the scope of the current work. Furthermore, because of the limitations of reliable corpus annotation, some important claims made in Chapter 3 have not been able to be tested here. Firstly, we claimed that focus projection rules such as those advanced by Selkirk (1995) are not necessary if we accept that a kontrast is the most prominent element in its scope; and that the scope of the theme/rheme unit containing that kontrast is therefore defined by prosodic phrasing. Following on from this, we claimed this phrasing is recursive, so prominence, and therefore kontrast interpretation, can be defined over several phrases. As we have said, this claim could not be directly tested since it was not feasible to annotate recursive phrasal structure; and we could not find a means of marking the scope of a theme/rheme unit independently of phrase or syntactic structure. Secondly, we have claimed that tonal pitch accent type is not important for the signalling of information structure, though nuclear accent shape may vary to express related affective and illocutionary connotations. It was decided not to annotate the corpus with ToBI pitch accent type, since we have already established that the crucial distinction, between L+H* and H*,

cannot be reliably defined nor made by annotators. Therefore we could not test our claim directly, though we once again found the expected prominence distinction between theme and rheme accents. However, we will look at both of these issues in the next chapter, which examines utterances from the conversation marked for theme/rheme status in detail to show more clearly the effect firstly of metrical structure on the perception of kontrast scope; and secondly, the effect of accent shape on interpretation.

Chapter 7

Illustrations from a Dialogue

In Chapter 3, we advanced our theory about how information structure is conveyed using prosodic prominence and phrasing. This theory involved several implications about the nature of both information structure and prosodic structure. In the last chapter, using statistical models, we were able to show that the distribution of information structural and prosodic properties in a wide-ranging corpus of spontaneous speech was consistent with our claims about the relationship between these two structures. However, because of the limitations of corpus annotation, we could not directly test whether an utterance with a given information structure would have the prosodic structure predicted by our theory, and vice versa. Therefore, in this chapter, we look at short extracts taken from the Switchboard dialogue used for the last experiment in the previous chapter (i.e. with full kontrast, information status and theme/rheme annotation), to test specific predictions about how information structure will be manifested prosodically. In particular, we look at how givenness, focus projection, *restricted* kontrast and theme/rheme status are signalled by prominence and phrasing. We finish by discussing the implication of our claim that theme/rheme status is signalled by prominence and not pitch accent type, by looking at the types of 'meanings' correlated with themehood that could have led to the intuition of a distinct 'theme accent' (or L+H*). Through our discussion in this Chapter, we find we are able to make advances on some of the open questions about both the nature of both prosody and information structure raised in Chapters 2 and 3.

The conversation is between A, a 28 year-old man from the Northern US, and B, a 51 year-old woman from the Western US. The general topic is the US federal budget, which at the time (1992), was in deficit. They begin talking about the prison systems, where B believes prisoners should have to work so the system is self-funding. They then move on to the education budget, where B argues that more money should be spent on in-work training schemes like apprenticeships, as in European countries, because college is not suitable for many high-school leavers. They finish with A offering his opinion that the deficit is caused

by unfair trading relationships between the US and other countries such as Japan. The sound files for all of the utterances discussed, along with Praat textgrids, can be found on the attached CD.[1]

7.1 Givenness

We began our discussion in section 2.2.2 with the often noted observation that new items tend to be accented, and given items deaccented. As we saw through the discussion in the rest of that chapter and the next, the accuracy of the observation depends a lot on what one means by given, and what one means by (de-)accented. In sections 2.2.2.1 and 2.2.2.2, we set out two distinct, but overlapping, notions about how to define givenness, i.e. *discourse givenness* and *relative givenness*. *Discourse givenness* describes a scale from the referent being completely new in the discourse, inferable from something said before or the speakers' mutual knowledge, to previously mentioned. *Relative givenness* describes whether the referent is given relative to the current proposition. As we discussed extensively in Chapter 3, the relevant prosodic relationship for both of these types of givenness is with relative prominence, not with (de-)accenting. That is, a referent is interpreted as relatively given in a proposition if it is less prominent, within metrical prosodic structure, than the surrounding elements. We suggested that, where they make different predictions, discourse givenness has less of an effect on relative prosodic prominence than relative givenness; although elements that are discourse given may tend to be less acoustically prominent overall. In section 2.2.2.1, we saw that, theoretically, relative givenness marking and kontrast marking are complementary. The results from the last chapter, i.e. that kontrastive elements align with nuclear prominence, support this position. However, we were not able to show clearly how relative givenness affects expected relative prominence structure. We were also not able to flesh out the interpretative differences between relative givenness and kontrast which we have previously noted. Additionally, we found only small effects of information status (i.e. discourse givenness) on prominence, although this could have been because the design of our experiments did not capture these effects well. Here, we demonstrate how these phenomena work using selected examples from our chosen dialogue, looking in detail at the effect on the acoustic correlates of prominence.

In (7.1), the participants are discussing how the federal education budget should be spent. B is saying that there should be alternatives to college, and A agrees that only a small number of people go to college and pass (see (7.7)):

[1]Please note these are also available at the following URL: `http://homepages.inf.ed.ac.uk/s0199920/thesis/`.

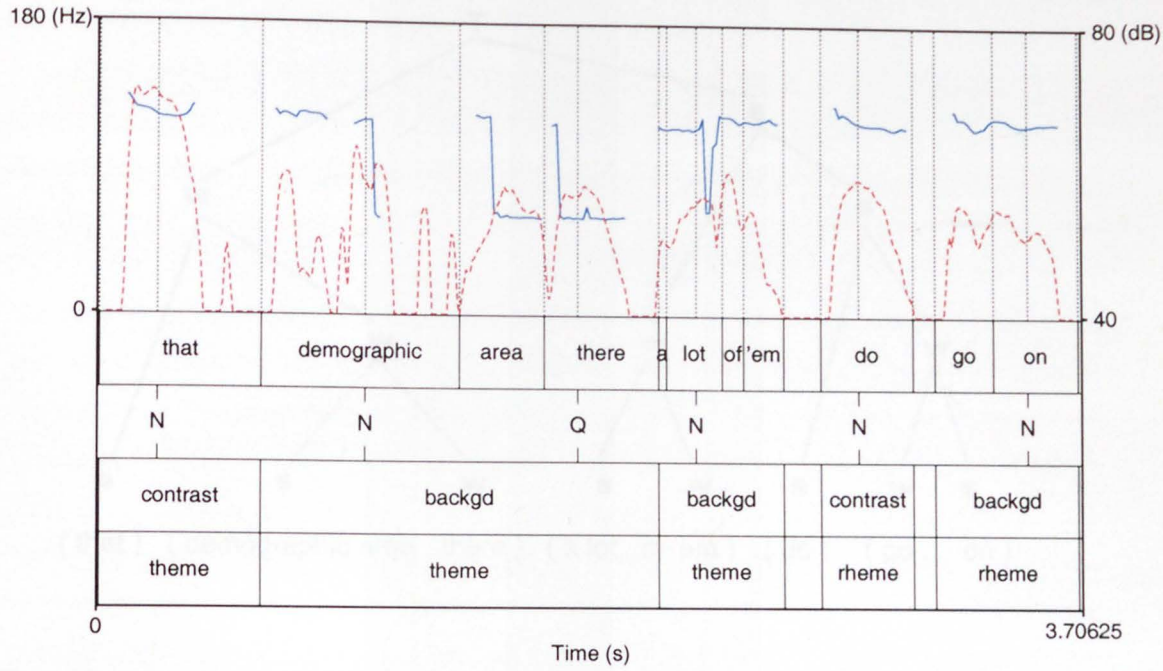


Figure 7.1: $f0$ trace (blue line) and intensity curve (dashed red line) for (7.1), along with the word transcript, accent type, kontrast status and theme/rheme annotation. Note theme/rheme annotation is per prosodic phrase.

(7.1) A: Most people I knew have gone and got their degree ...

B: ... Yeah, but there's a lot of them out there that haven't ...

A: ... it was basic ... middle class, upper middle class area ...

A: **THAT demographic area there, a lot of 'em DO go on**

Figure 7.1 shows a representation of the acoustic properties of the utterance, along with the kontrast and theme/rheme annotation (the sound file is *demographic*). Kontrast is marked by annotated kontrast type (e.g. *contrast*, *subset*, rather than [±kontrast]). Theme/rheme status is marked per prosodic phrase. The first two phrases are the theme, and the last three the rheme (we will return to theme/ rheme derivation below).[2] In the theme phrase, everything is relatively given compared to *that. demographic area* is inferable from *middle class area* (which is used in relation to the ability to go to and pass college); *there* probably serves to

[2]Note that *a lot of 'em* could also plausibly be part of the theme, although in the author's judgement this phrase sounded like it was more closely grouped with the rheme. As we noted in section 2.2.3.1, there is a general indeterminacy between the background of the theme and the background of the rheme. This does not materially affect the argument being made here.

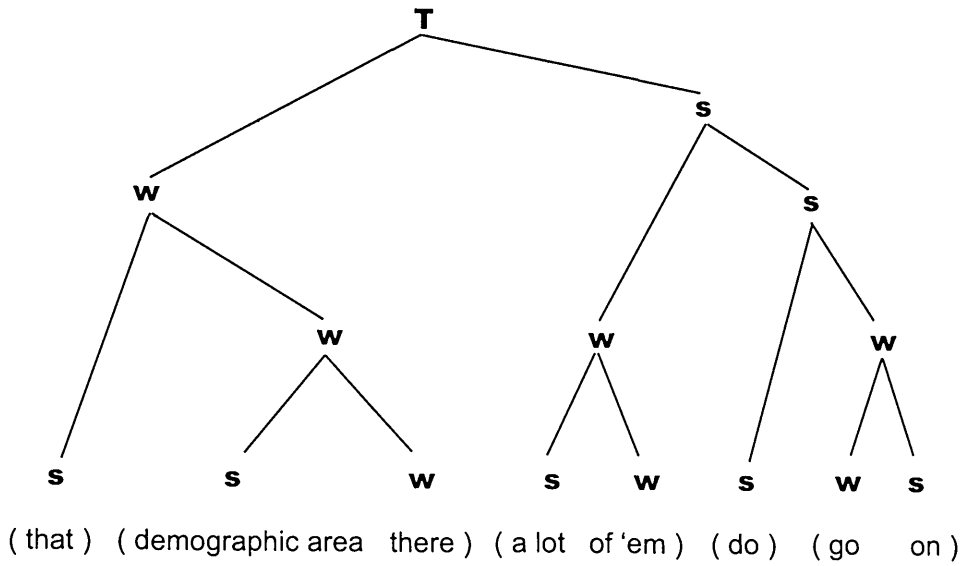


Figure 7.2: Possible metrical structure showing prominence relationships between words in (7.1). Phrase boundaries are shown by parentheses.

highlight that this is the relevant reference. The demonstrative adjective *that* is not-given in relation to these, probably intended to contrast with an alternative set of *other demographic* areas. Similarly in the rheme phrase, *go on* is relatively given, since the discussion is about *going on to college*. *a lot of 'em* is also relatively given since A has just mentioned his middle class friends getting degrees; although *lot* may be intended to contrast with the *lot ... that haven't* (this wasn't marked by the annotator). *do* is not-given, intended to form an alternative set with its polar *do not*.

We can see in Figure 7.2 how these relative prominence relationships are directly reflected in the branching structure of the prosodic tree. *demographic* is strong relative to *there*, since *there* is semantically 'light'. *that* is strong relative to *demographic* as it is the kontrast of the theme. Similarly, *on* is strong relative to *go*, since this is the default. *lot* is strong relative to *'em* since it is less relatively given. *do* is strong relative to *on* and *lot*, since *do* is kontrastive. Finally, *do* is strong relative to *that*, showing that *do* is the rheme, and *that* the theme.

What is interesting is how this structure is manifested in the acoustic properties of the utterance. Evidently, there is no relationship between givenness and accenting per se. *demo-*

graphic, *lot* and *on* all carry nuclear accents (the latter two since they are the only accent in their phrase); even though, by most accounts, they are relatively and discourse given. This is partly because A is speaking slowly and pausing often, however, the information structure is still clear. The perception of kontrast status comes from association with nuclear prominence over multiple phrases: the second phrase is subordinate to the first, so *that* is nuclear for the theme; and the first and third rheme phrases are subordinate to the second, so *do* is nuclear for the rheme. In fact, there is a sense in which we would not want to call the accents on *lot* and *on* 'nuclear' at all, as they do not seem like 'perceptual centres' in the same way as the accents on *that* and *do* (and to a lesser extent *demographic*). As we discussed in section 3.1.2, we expect such mismatches between the perception of phrasing structure and prominence structure when information units span over multiple prosodic phrases. Further, these subordination relationships are not primarily conveyed in terms of variations in pitch. There are no clear pitch movements to indicate accents; and while the subordinate phrases in both the theme and rheme do have lower pitch overall, the effect is very slight. Rather, *that* and *do* are made prominent by increased intensity and duration, both being much louder and relatively longer than the words in their respective subordinate phrases. The phrase *do* is also separated by pauses. This is a trend seen frequently in Switchboard conversations, particularly with male speakers. It shows once more that the relevant property to convey basic semantic elements (i.e. information structure) prosodically is relative prominence, not pitch accent type.

There are many examples in our chosen conversation of the strong relationship between relative givenness and relative prominence. We have space for only one more. The second part of this example will also show why the relationship between relative givenness and relative prominence is not direct, but is subject to other interacting constraints. In (7.2), B is arguing that people should be educated on the job, rather than at college, and A is countering that this would be too expensive for many businesses:

(7.2) A: One of the biggest things now is paralegals... trying to get more people in that field. But they can't just bring somebody in, without even having been to school in that area...

B: Excuse me I see it being done...

B: **I know a friend that works for MY lawyer, that has NO training whatsoever, and she's TRAINING her**

Figure 7.3 shows the acoustic representation along with the phrasal, kontrast and information status annotation for the utterance (sound file *lawyer*). We can see that the prominence relationships in the first part of the utterance can be derived, as before, from the in-

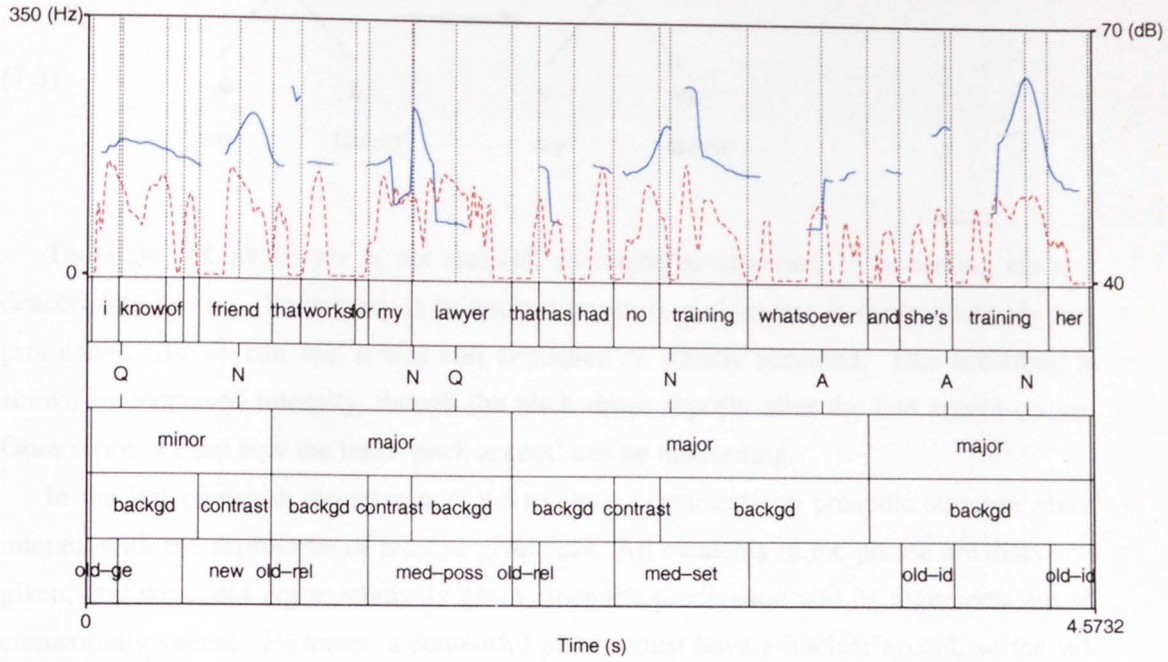
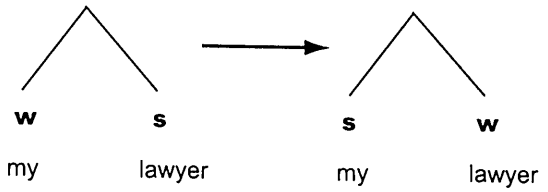


Figure 7.3: $f0$ trace (blue line) and intensity curve (dashed red line) for (7.2), along with the word transcript, accent type, phrase type, kontrast status and information status annotation.

formation structure. *friend* is relatively given compared to *works for my lawyer*. In terms of the relevant properties of the trainee (i.e. salient alternative sets), the property of them working for B's lawyer is more important than them being a friend. The subordinate relationship between *friend* and *my* is shown, as in the last example, by the ordering of equally high nuclear accent peaks. The stress on *works for my lawyer* would normally be on *lawyer*. However, *lawyer* is inferable from the topic of *paralegals*. Therefore *lawyer* is 'deaccented' (in the sense of Ladd 1980). By metrical reversal, the nuclear accent occurs on *my*, as shown below. Note that this could equally be interpreted as a kontrast on *my*, meant to contrast with *other lawyers* that A is talking about. As discussed in section 2.2.2.1, the implications in alternative semantics of a referent being 'not relatively given' and kontrastive are the same. However, in some cases one explanation works better than the other. We return to this in a later example, and suggest that the interpretative differences are related to the salience of properties of the alternative set, which in turn are cued by increasing degrees of relative prominence.

(7.3)



The status of *my lawyer* is not realised, as might be expected, by accenting *my* and deaccenting *lawyer*. Rather, *my* is in nuclear position, and *lawyer* is made relatively less prominent. As we can see, it was still annotated as weakly accented. This accenting is shown by increased intensity, though the pitch drops sharply after the late accent on *my*. Once more, we see how the term 'pitch accent' can be misleading.

In the last phrase in the utterance, we see how constraints on prosodic structure itself interact with the expression of relative givenness. All elements in the phrase are discourse given, and we could argue relatively given since the predication and its arguments are all contextually salient. However, a contentful phrase must have a nuclear accent, so the following structure is indicated: the theme *she* is kontrastive, i.e. opposed to other lawyers, but because it is pronominal and repeated, it does not have enough weight to form a phrase, so it is just accented. Further, a sole nuclear accent on *she* would change the meaning of the phrase, which is a kontrast on the polarity of the predication, i.e. opposed to *not training her*. This structure would normally lead to 'default' prominence on the object, but here *her* is pronominal and short, so *training* has to be nuclear, i.e.:

```
                *
        *       *
        *       *       *
(7.4)   she's=↑ training  her=↓
```

One possible test for this could be to add an adverb such as *now* or *voluntarily*. In the author's judgement at least, this would attract nuclear prominence, showing the relative givenness of the other elements in the phrase. In any case, this example shows how different constraints on prosodic and information structure interact.

It is difficult to test whether discourse givenness has an overall effect on prominence on this data. This is largely because of the nature of the information status coding, which includes a big proportion of mediated entities, where it is unclear where on the scale from given to new the referent is predicted to lie. Further, discourse givenness and relative givenness do in fact often make the same predictions regarding relative prominence. For instance, in the last utterance, all the *old-identity* words except *she's* were not accented, and the one *new* word was nuclear accented. However, when the two conflict, relative givenness/kontrast is a

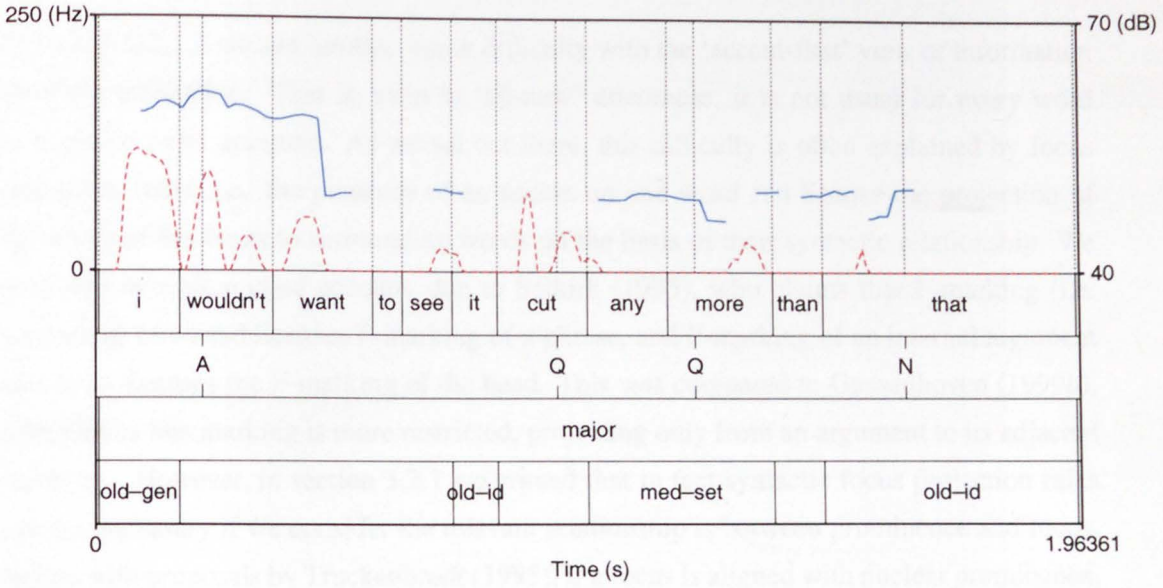


Figure 7.4: $f0$ trace (blue line) and intensity curve (dashed red line) for (7.5), along with the word transcript, accent type, phrase type and information status annotation.

better predictor of relative prominence, as we saw with the status of the mediated *my lawyer* and *she's* in the last utterance.

On the other hand, we could find a few instances where downstepping and discourse givenness are related (cf. Baumann 2005). In (7.5), the speakers are generally talking about the US budget, but had been speaking about the prison systems when B changes the topic:

(7.5) B: As far my **DEFENCE** budget... they're cutting it back now what 25%

B: **I wouldn't want to see it cut any more than THAT**

As expected, *I* and *it* are unaccented as *old* (see Figure 7.4, sound file *cutmore*). We can see that *that*, referring to the 25%, is perceived as nuclear. However, it is said with heavily reduced pitch and intensity, the nuclear perception probably coming from a combination of phrase position and lengthening. This is consistent with *that* being kontrastive (i.e. as opposed to a *more cuts*), as well as downstepped because it is highly accessible in the discourse.

7.2 Focus Projection and Pre-Nuclear Accents

In section 2.2.1.2 we saw another major difficulty with the 'accent-first' view of information structure realisation. That is, even in 'all-new' utterances, it is not usual for every word in a phrase to be accented. As we set out there, this difficulty is often explained by focus projection rules, i.e. the presence of an accent on one word can license the projection of the scope of the focus to surrounding words on the basis of their syntactic relationship. We reviewed one often cited account, due to Selkirk (1995), who claims that F-marking (i.e. accenting) of a head licenses F-marking of a phrase, and F-marking of an internal argument of a head licenses the F-marking of the head. This was compared to Gussenhoven (1999*b*), who claims this marking is more restricted, projecting only from an argument to its adjacent predicate. However, in section 3.2.1 we argued that in fact syntactic focus projection rules are not necessary if we consider the relevant relationship is between prominence and focus. In line with proposals by Truckenbrodt (1995), if a focus is aligned with nuclear prominence, it is automatically interpreted as having scope over all the material within the scope of that prominence, defined by prosodic phrasing structure. We set out evidence from Büring (to appear) showing traditional focus projection rules are both too restrictive and unconstrained to account for even simple examples; whereas under the prominence view these and more complex examples immediately receive a straight-forward explanation. We developed this idea to claim, as we saw in the last section, that the relative prominence relationship can hold between nuclear accents across phrases; and that the theme-rheme relationship is part of the same story, i.e. it is weak-strong. One consequence of this is that other accents generally appear as required by metrical structure and are not directly meaningful. However, this is not always true, as reversal of the expected weak-strong relationship in the pre-nuclear domain, and particularly strong pre-nuclear accents, can signal kontrastive readings. The results from the last chapter generally support these ideas. However, we were not able to show precisely how the metrical view leads to better predictions than the syntactic view, and so will address this here. We did see that increased acoustic prominence, compared to expected prominence, makes a kontrastive interpretation more likely; here we will briefly show an example where this occurs.

(7.6) comes at the beginning of the discussion in (7.1) and (7.2), where B is arguing more of the education budget should be spent on on-the-job training, and less on college:

(7.6) B: But see, we don't even push the fact, to the high school kids, that there's other means of education out there rather than college... to go either as an apprentice, which they do in other countries...

B: **why not APPRENTICE out to a, a COMPANY and learn from down on**

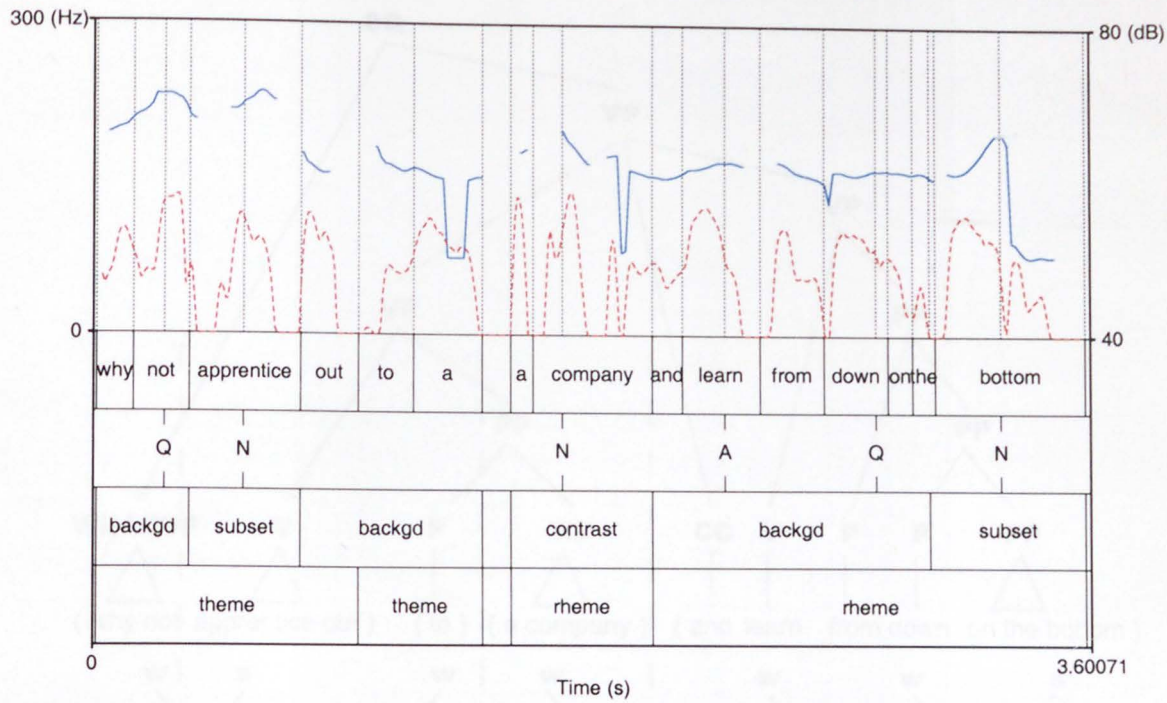


Figure 7.5: $f0$ trace (blue line) and intensity curve (dashed red line) for (7.6), along with the word transcript, accent type, kontrast status and theme/rheme annotation. Note theme/rheme annotation is per prosodic phrase.

the BOTTOM

Figure 7.5 shows the acoustic representation, along with the kontrast and theme/ rheme annotation as before (sound file *company*). We can see that the first two phrases are thematic, with *apprentice* as the head. *apprentice* is given, having just been mentioned, however, here it is meant to be kontrastive, i.e. as opposed to *college*. The rheme phrase is *a company and learn from down on the bottom*, broadly describing what the *kid* is supposed to gain from the *apprenticeship*. We can see straight away that the theme/rheme division does not match syntactic phrasing (at least according to traditional analysis). Figure 7.6 shows the relationship between the syntactic and prosodic structure of the phrase.[3] Firstly, the specifier of the top SQ node and the verb are closely paired in the prosody, though they have three layers of intervening structure in the syntax. Both the higher co-ordinated VP, and the lower VP *apprentice out* are split between the theme and the rheme. Finally, the rheme phrase groups the object NP from the first VP with the second co-ordinated VP.

[3]Syntax structure is taken from the Penn treebank analysis, except that the top node structure is abbreviated,

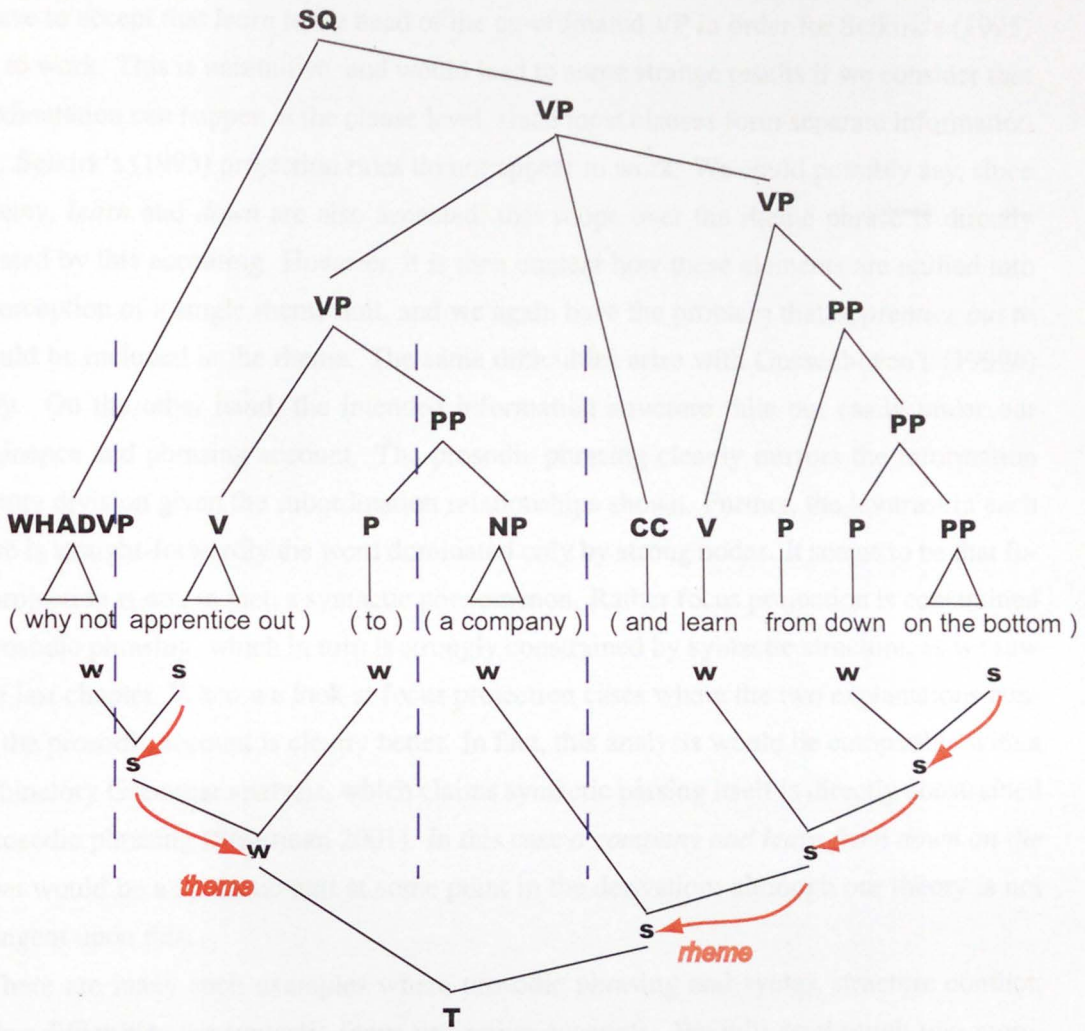


Figure 7.6: Relationship between syntactic structure and a possible metrical structure showing the relative prominence of words in (7.6). Arrows show how theme/rheme status is projected from the nuclear accent in the metrical structure. Dashed lines show mismatches between syntactic and metrical structure which cause problems for syntactic focus projection accounts.

These mismatches cause much difficulty for any syntactic projection account. If we accept that the scope of the rheme can project from *bottom* to the higher co-ordinated VP; there is no reason why *apprentice out to a* would not then be included in the rheme, unless

i.e. SBARQ and SQ are collapsed.

we appeal to phrasing. Further, projecting above the co-ordination is problematic in itself: we have to accept that *learn* is the head of the co-ordinated VP in order for Selkirk's (1995) rules to work. This is unintuitive, and would lead to some strange results if we consider that co-ordinatation can happen at the clause level, since most clauses form separate information units. Selkirk's (1995) projection rules do not appear to work. We could possibly say, since *company*, *learn* and *down* are also accented, that scope over the rheme phrase is directly indicated by this accenting. However, it is then unclear how these elements are unified into the perception of a single rheme unit, and we again have the problem that *apprentice out to a* would be included in the rheme. The same difficulties arise with Gussenhoven's (1999*b*) theory. On the other hand, the intended information structure falls out easily under our prominence and phrasing account. The prosodic phrasing cleanly mirrors the information structure division given the subordination relationships shown. Further, the kontrast in each phrase is straight-forwardly the word dominated only by strong nodes. It seems to be that focus projection is not, in fact, a syntactic phenomenon. Rather focus projection is constrained by prosodic phrasing, which in turn is strongly constrained by syntactic structure, as we saw in the last chapter. When we look at focus projection cases where the two explanations conflict, the prosodic account is clearly better. In fact, this analysis would be compatible with a Combinatory Grammar analysis, which claims syntactic parsing itself is directly constrained by prosodic phrasing (Steedman 2001). In this case *a company and learn from down on the bottom* would be a syntactic unit at some point in the derivation; although our theory is not contingent upon this.

There are many such examples where prosodic phrasing and syntax structure conflict, causing difficulties for syntactic focus projection accounts. We will go through two more, the first shows how kontrast marking interacts with relative givenness and the problems this causes for syntactic projection accounts. The second how the giveness/prominence account can be mediated by constraints on prosodic structure. (7.7) comes at the beginning of the extract in (7.1), discussing why college doesn't benefit many kids coming out of high school:

(7.7) A: I just read something the other day... only 60% or 40% go to college, and then, out of that percentage, only so many can get their degree...

A: But... **most people I KNEW, have GONE and got their degree**

The information structure indicated by the prosody is shown in (7.8a) and (7.9a) below (see Appendix G for the acoustic display, and kontrast and theme/rheme annotation, sound file *degree*). Each theme (θ) and rheme (ρ) unit forms its own phrase, and the F-marked word in each phrase falls on the nuclear accent:

(7.8) a. ([most people [I [knew$_F$]]])$_0$

b. [most people [I [knew$_F$]$_{VP}$]$_{CP}$]$_{NP}$

c. [most people [[I [knew$_F$]$_{VP}$]$_{CP}$]$_{NP}$]$_{Foc}$

(7.9) a. ([have [[gone$_F$] and [got [their degree]]]])$_\rho$

b. [have [[gone$_F$]$_{VP}$ and [got [their degree]$_{NP}$]$_{VP}$]$_{VP}$]$_{AuxP}$

c. * [[have [[gone$_F$]$_{VP}$ and [got [their degree]$_{NP}$]$_{VP}$]$_{VP}$]$_{AuxP}$]$_{Foc}$

As we can see in (7.8b) and (7.8c), for the theme phrase, the projection of focus to the whole NP could be equally well accounted for in Selkirk's (1995) scheme: *knew* is the head of the CP *I knew*, licensing projection to the CP; and *I knew* is the internal argument of *most people*, licensing Foc marking on the whole NP. However, while the prosodic account also covers the rheme phrase straight-forwardly, the syntactic account cannot (see (7.9b) and (7.9c)). *got their degree* is made less prominent, since it is repeated, so the nuclear accent falls on *gone*. The word *gone* has much higher pitch and intensity than the other words in the phrase. While in the last example, we might have accepted that F-marking of the last word in a co-ordinated VP might, through 'default' prominence, license marking for the whole VP; in this case the F-marking is on the left branch of the co-ordination, so 'default' marking cannot apply.

Our last example shows interaction with other constraints on structure. In this extract (which follows on from (7.25)), the speakers have been talking about the education budget, when B changes the topic to ask A what he would do about the deficit (we discuss the *Japan* sentence in (7.30) below):

(7.10) A: the deficit basically is the trade surplus between the other countries... we have... more money going out, and too many goods coming into this country... part of that problem... is... like **JAPAN still does not let us COMPETE FAIRLY in their COUNTRY**, and obviously

A: **the demand for their goods is quite HIGH here, so they can get their GOODS in here**

Figure 7.7 shows the acoustic representation, along with the information status and phrasal annotation (sound file *goods*). We will concentrate on the last phrase, where the nuclear accent on *goods* licenses focus over the whole phrase. All the elements in the phrase are given since they are contextually salient, having been used in similar propositional contexts in the immediately preceding discourse. Therefore, broadly, the kontrast is on the

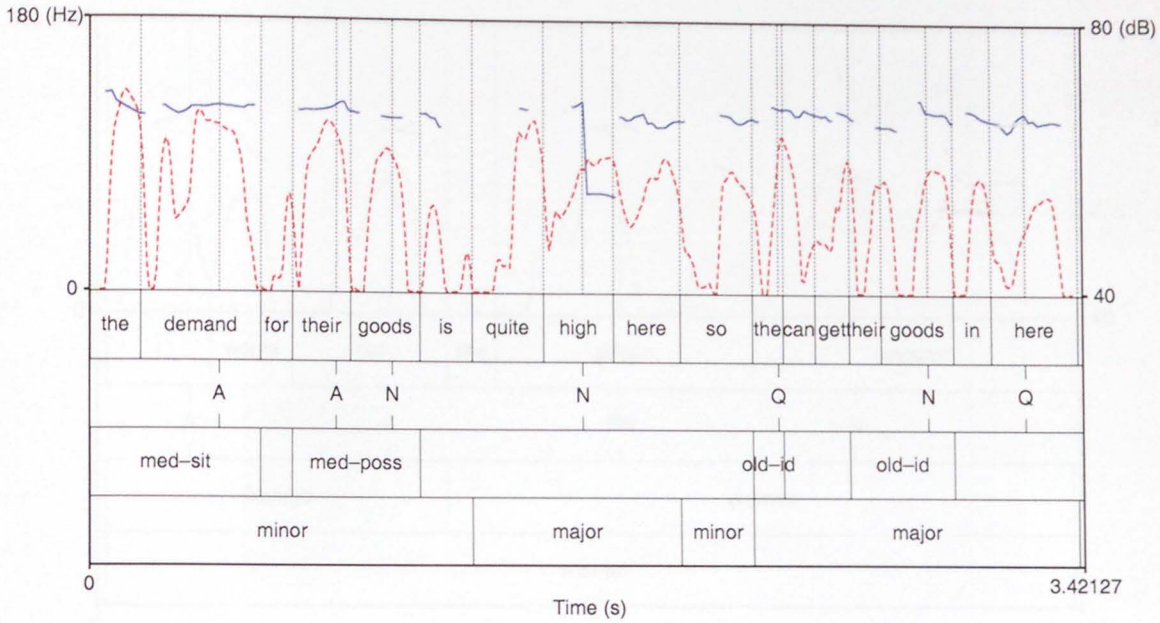


Figure 7.7: $f0$ trace (blue line) and intensity curve (dashed red line) for (7.10), along with the word transcript, accent type, information status and phrase type annotation.

polarity of the clause, i.e. as opposed to *not getting their goods in here*. The question is why the nuclear accent is not on *in*, being the 'least' given element (since it wasn't mentioned in the main clause). The answer is that this would change the scope of the focus, and therefore, subtly, the meaning. This is because we do not expect a preposition to be accented; and so if it is, this implies narrow focus on it, i.e. as opposed to something like *out of here*, which is odd in the context. Similarly, a nuclear accent on *here* would introduce an unintended salient alternative set of *other places*, so it has to occur on *goods*, as follows:

(7.11) they can get their=↓ goods in=↓ here=↓

(7.12) * they can get their goods IN here

This introduces the general problem which we have been skirting in the discussion up till now; and that is whether there is a consistent prosodic distinction between those parts of a kontrast which are not nuclear, and the backgrounded parts of a theme or rheme phrase. This is essentially the same question as whether there is a consistent distinction between broad and narrow focus. It is difficult to test this directly in this corpus, as there are very

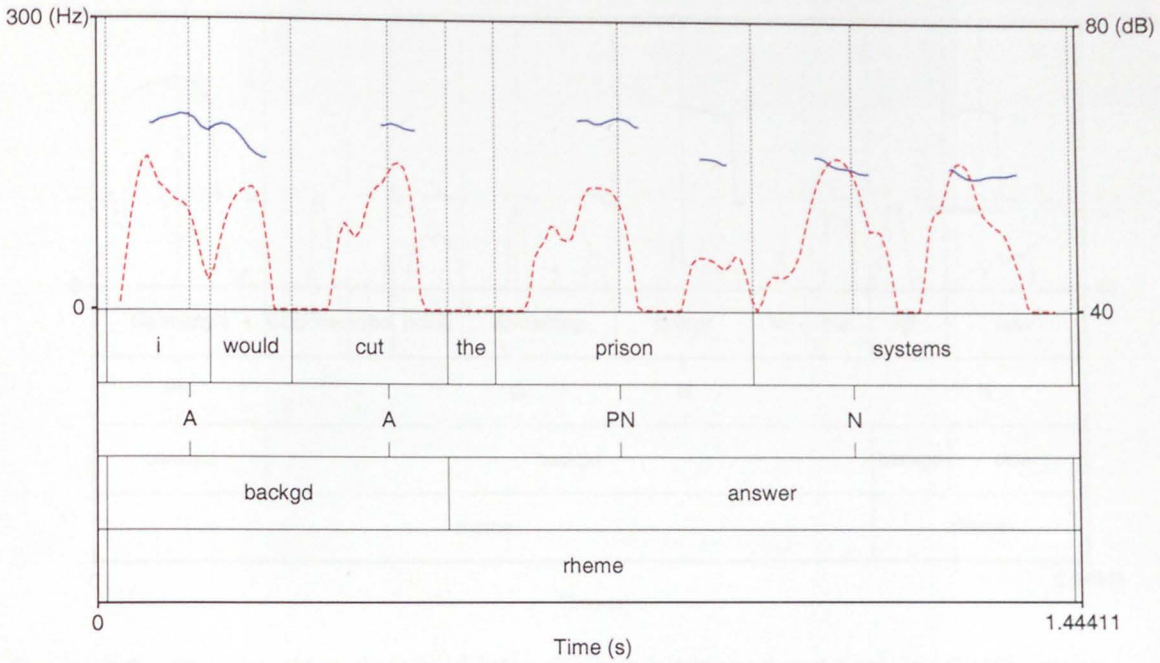


Figure 7.8: $f0$ trace (blue line) and intensity curve (dashed red line) for (7.13), along with the word transcript, accent type, kontrast status and theme/rheme annotation. Note theme/rheme annotation is per prosodic phrase.

few examples of the usual broad/narrow focus paradigm, i.e. question-answer pairs. This is because of the genre of the corpus, i.e. conversations on general topics between strangers, where there are very few direct questions asked. However, we find this one in our dialogue, at the start of the conversation:

(7.13) B: my first comments on the budget...

A: ... what would be the first thing you'd cut? defence?

B: Surprisingly, no

B: **I would cut the PRISON SYSTEMS**, and let them self-support

B's answer shows the classic division into background and focus, even the same words are repeated (cf. section 2.2.1.1):

(7.14) [I would cut]$_{bkgd}$ [the prison systems]$_{whFoc}$

However, as can be seen in Figure 7.8, this does not lead to complete deaccenting of the background, as might be expected: there are clear accents on *I* and *cut* (sound file *prison*).

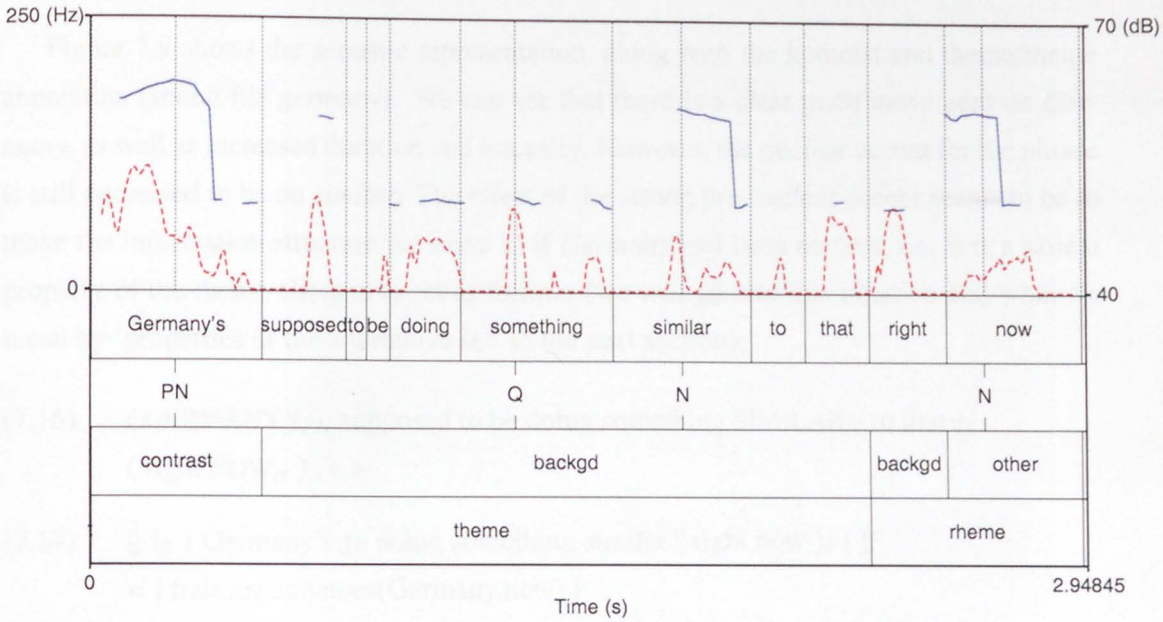


Figure 7.9: $f0$ trace (blue line) and intensity curve (dashed red line) for (7.15), along with the word transcript, accent type, kontrast status and theme/rheme annotation. Note theme/rheme annotation is per prosodic phrase.

Predictions of most focus-to-accent theories are clearly violated. On the other hand, the accents on both *prison* and *systems* are particularly strong. The phrase is treated as a compound noun, so the nuclear accent is on *prison*. This has the highest pitch and is lengthened. The accent on *systems* is marked by lengthening and the greatest intensity, though pitch is lowered in post-nuclear position. Therefore, the narrow focus seems to be marked by increased prominence on the answer, not by deaccenting the background.

Looking at the other side of the story, while in all the examples so far kontrastive elements have been nuclear on at least one level of phrasing, particularly strong pre-nuclear accents (PN) do seem to trigger a kontrast as well. (7.15) continues from the discussion in (7.2), where B was arguing that more of the education budget should be spent on apprentice schemes, like in other countries. A responds:

(7.15) A: maybe [the government] needs to help promote that more, give incentives... like they will pay some of [the company's] costs, or give tax breaks for them [the company] to train people...

B: ...**GERMANY'S supposed to be doing something SIMILAR to that right**

NOW

Figure 7.9 shows the acoustic representation, along with the kontrast and theme/rheme annotation (sound file *germany*). We can see that there is a clear pitch movement on *Germany*, as well as increased duration and intensity. However, the nuclear accent for the phrase is still perceived to be on *similar*. The effect of the strong pre-nuclear accent seems to be to make the information structure the same as if *Germany* had been nuclear, i.e. it is a salient property of the theme alternative set as follows (we will go into this notation and what we mean by 'properties of the alternative set' in the next section):

(7.16) (GERMANY's$_{PN}$ supposed to be doing something SIMILAR$_N$ to that)$_\theta$
(right NOW$_N$)$_\rho$ =>

(7.17) [[[$_S$ [Germany's]$_F$ doing something similar [right now]$_F$]]]^o
= { training_schemes(Germany,now) }

(7.18) [[[[Germany's]$_F$ doing something similar]]]$^{f\theta}$
= { training_schemes(x) | $x \in E$ }, where E is the domain of countries

(7.19) [[[Germany's doing something similar [right now]$_F$]]]$^{f\rho}$
= { training_schemes(Germany,x) | $x \in E$ }, where E is the domain of time

It is uncertain how these 'PN' accents should be incorporated in our representation of metrical structure. We could hypothesise that the effect of the PN is to add a level to the accent in the metrical grid, so that effectively there will be two nuclear accents in the phrase. This would lend them a categorical status, however, and it is still unclear whether this is justified.

From our examples so far and the results in the last chapter, the answer to our background marking question seems to be "sometimes". In general, the more relatively prominent an element is, the more likely it will be taken as a salient property of the alternative set; and therefore, in comparison, the more likely the surrounding elements are to be taken as backgrounded. Broadly, our contention holds that most non-nuclear accents are not directly meaningful. However, their prominence can indicate whether they form part of the salient properties of the rheme or theme alternative set. Further, when the expected prominence relation is preserved, the interpretation may be ambiguous (e.g. the interpretation of *friends* in (7.2)). When the intended division is obvious from the context, it may not be marked by increased prominence.

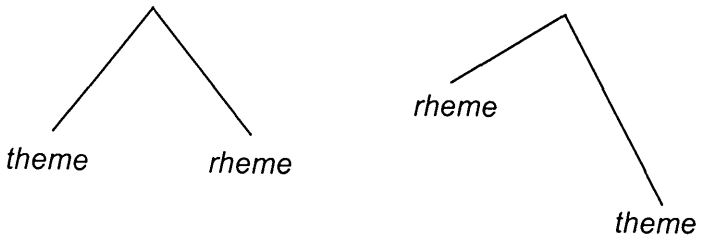


Figure 7.10: Diagrammatic representation of the signalling of the relative metrical prominence of theme and rheme nuclear accents.

7.3 *Restricted* Kontrast and Theme/Rheme Structure

One of the central claims in this thesis has been that themes are distinguished from rhemes by prosodic subordination, not pitch accent type. This is coupled with a claim about how prosodic subordination is manifested in metrical prosodic structures, i.e. by order when the elements are weak-strong, and by acoustic prominence when they are strong-weak (see Figure 7.10). In section 2.2.3 we argued that one of the reasons themes are often thought to have distinct accents is because of the confounding of two semantic dimensions: that is, since most themes are not kontrastive, they are not distinguished from the background in a rheme. Therefore, when they are kontrastive, they are often particularly emphasised, giving a *restricted* kontrast interpretation, i.e. a salient, and restricted, alternative set. The results in Chapter 4 supported this claim, showing no consistent difference in pitch accent type, but a distinction in pitch accent height, between theme and rheme accents. This was upheld in the small corpus study at the end of the last chapter. Here we show, using examples from our dialogue, how relative prominence across phrases leads to theme/rheme interpretation; and in particular, how prominence patterns within utterances generate theme and rheme alternative set presuppositions for that utterance. We will see how increased prominence, and the context, can lead to a *restricted* kontrast interpretation. And we will show that accent shape does not seem to play an important role.

(7.20) follows on from B's comment in (7.13). A suggests that what B means is that private enterprise should run the prisons:

(7.20) A: they're talking about having it [the prison system] as a business... so...

A: **the GOVERNMENT doesn't have to DEAL with it**

Figure 7.11 shows the acoustic representation, along with the kontrast and theme/rheme

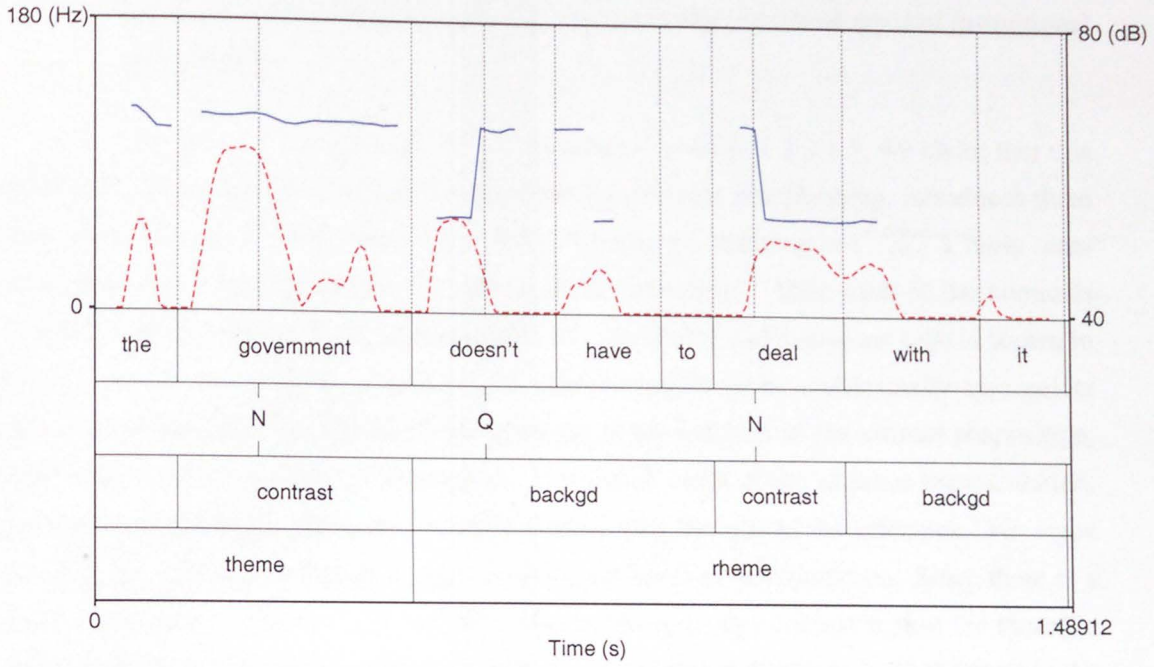


Figure 7.11: $f0$ trace (blue line) and intensity curve (dashed red line) for (7.20), along with the word transcript, accent type, kontrast status and theme/rheme annotation. Note theme/rheme annotation is per prosodic phrase.

annotation (sound file *govdeal*). As in previous examples, the nuclear accent on *government* marks it as the head of its phrase and a kontrast. The nuclear accent on *deal* marks it as the head of its phrase and a kontrast. Theme/rheme status is determined by the ordering of the phrases. The rheme accent has lower pitch and intensity than the theme accent, but this is not sufficient to reverse the expected prominence relationship between the phrases. Therefore the first phrase is the theme and the second the rheme, as in (7.21).

(7.21) (the GOVERNMENT$_N$)$_\theta$ (doesn't have to DEAL$_N$ with it)$_\rho$ =>

(7.22) [[[$_S$ [the government]$_F$ doesn't have [to deal]$_F$ with it]]]^o
= { not_deal_with(government,prisons) }

(7.23) [[[[the government]$_F$ doesn't have to deal with it]]]$^{f\theta}$
= { not_deal_with(x, prisons) | $x \in E$ }, where E is the domain of institutions & *business* $\in E$

(7.24) [[[the government doesn't have [to deal]$_F$ with it]]]$^{f\rho}$
= { P(government,prisons) | $P \in E$ }, where E is the domain of types of institutional relationship

Using the notation from Rooth (1992) introduced in section 2.2.1.3, we claim that this information structure, derived directly from the prominence and phrasing, introduces three separate meanings. Firstly, there is the ordinary semantic meaning in (7.22), a fairly standard derivation of the propositional content of the utterance.[4] Then each of the kontrasts introduces a presupposition of an alternative set. The theme alternative set ($f\theta$) is shown in (7.23).[5] As we discussed in section 2.2.1.3, this alternative set is a contextually appropriate set of other elements that could fill the position of the kontrast in the current proposition, here defined as the *domain of institutions*. If we look again at the acoustic representation, *government* has higher pitch, and is much louder, than the rest of the utterance. We argue that this increased prominence triggers a *restricted* kontrast interpretation. Since there is a salient alternative available, i.e. *business*, this marking in this context makes the thematic alternative set highly likely to be restricted to { *government*, *business* }. In the rheme, the verb is kontrastive, its alternative set presupposition ($f\rho$) is shown in (7.24). This is harder to represent using the notation here, so we will not try to formalise this completely, but use P to represent an element in a contextually appropriate set of predicates. *deal* is not especially prominent, and there are no salient alternatives in the context, so a *restricted* kontrast interpretation is not triggered. We can see quite clearly that there is no characteristic distinction in pitch accent shape between theme and rheme. Indeed A does not seem to vary his intonation very much at all, with the pitch contour for many utterances being completely flat; but he still manages to convey information structure unproblematically.

The next example is similar, except that we see the effect when the scope of the theme and rheme spans multiple phrases. (7.25) comes before the extract in (7.10), where our speakers were talking about education when B changes the subject:

(7.25) B: you threw that question on me about the deficit... what would you do?

A: my perception of the budget... the goverment... has so much money to spend, and there's not enough money to spread around

A: **but the DEFICIT basically is the TRADE surplus between the other COUNTRIES**

[4]For simplicity's sake we consider the negation to be part of the predication, though in a full derivation this would be a separate operator.

[5]For ease of exposition we show the predication filled. However, since the predication is also the rheme kontrast, this should be a contextually appropriate predicative relationship between an alternative set of *institutions*, and *prisons*.

As we can see in (7.26), *deficit* is the kontrast in the theme, which has scope over the first two prosodic phrases. This is indicated by a large pitch movement and high intensity on *deficit* and a much lower and less intense nuclear accent on *basically* (see Appendix G for the acoustic representation, and the information status and theme/rheme annotation, sound file is *surplus*). The rheme has a scope over the last two phrases, with nuclear accents on *trade* and *countries*.

(7.26) ((the DEFICIT$_N$) (basically is))$_\theta$
((the TRADE$_N$ surplus between) (the other COUNTRIES$_N$))$_\rho$ =>

(7.27) [[[$_S$ [the deficit]$_F$ is [the trade surplus between the other countries]$_F$]]]^o
= { is(deficit,trade_surplus_with_other_countries) }

(7.28) [[[[deficit]$_F$ is the trade surplus between the other countries]]]$^{f\theta}$
= { is(x,trade_surplus_with_other_countries) | $x \in E$ }, where E is the domain of US economic measures & *the budget* $\in E$

(7.29) [[[the deficit is [the trade surplus between the other countries]$_F$]]]$^{f\rho}$
= { is(deficit,x) | $x \in E$ }, where E is the domain of things that affect the economy

As before, the ordinary semantic meaning of the utterance is shown in (7.27). For ease of exposition we represent *trade surplus with other countries* as a single argument, though of course in a full derivation this would need to be broken up into separate, or nested, predications. The theme alternative set is shown in (7.28). Once more, the particularly strong accent on *deficit*, together with the contextual availability of an alternative, lead to a restricted alternative set of { *deficit*, *budget* }. The rheme alternative set is shown in (7.29). Here we can begin to see more clearly what we mean by prominence affecting the relevant properties of the alternative set. The nuclear prominence for the whole rheme phrase is on *countries*. However, there is also a strong nuclear accent on *trade*. The effect of this is to make the salient alternatives to the rheme those that contrast with *trade*, e.g. as opposed to the *fiscal surplus*, as well as *other countries*, e.g. as opposed to *the US*; though of course the latter is ambiguous since it falls in the default prominence position.

We can see this idea more clearly in the next example, which we saw in the extract in (7.10). As shown in (7.30), the first phrase is the theme and *Japan* is its kontrast. The rheme comprises the last three phrases, headed by *compete*, *fairly* and *country* respectively.

(7.30) (JAPAN$_N$ still)$_\theta$
((does not let us COMPETE$_N$) (FAIRLY$_N$) (in their COUNTRY$_N$))$_\rho$

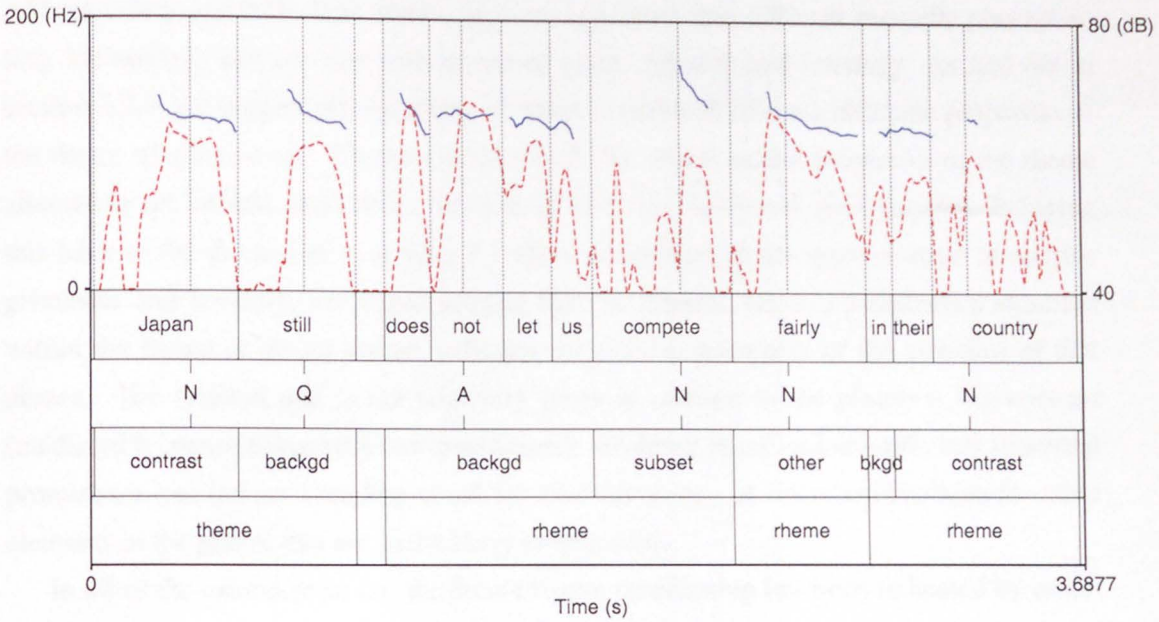


Figure 7.12: $f0$ trace (blue line) and intensity curve (dashed red line) for (7.30), along with the word transcript, accent type, kontrast status and theme/rheme annotation. Note theme/rheme annotation is per prosodic phrase.

(7.31) [[[$_S$ [Japan]$_F$ still does not let us [compete fairly in their country]$_F$]]]^o
= { let_compete_fairly_in_their_country(Japan,US) }

(7.32) [[[[Japan]$_F$ still does not let us compete fairly in their country]]]$^{f\theta}$
= { let_compete_fairly_in_their_country(x,US) | $x \in E$ }, where E is the domain of countries

(7.33) [[[Japan still does not let us [compete fairly in their country]$_F$]]]$^{f\rho}$
= { P(Japan,US) | $P \in E$ }, where E is the domain of economic relationships

(7.34) { P } = { govern fairly in their country,
compete aggressively in their country,
compete fairly with other countries ... }

As in the last examples, this prosodic structure leads to the ordinary semantic meaning in (7.31), and the theme alternative set in (7.32). Our interpretation of the rheme alternative set in (7.33) is mediated by the prominences within the rheme phrase. As we can see in Figure 7.12, because of its position, *country* is the nuclear accent of the whole rheme phrase;

but *compete* and *fairly* are both nevertheless made particularly prominent (sound file *japan*). Although they are both short words, they are separated into different prosodic phrases so they are nuclear, and are said with increased pitch, duration and intensity. As laid out in section 3.2.3, we suggest that the effect of this is to make all of these elements properties of the rheme alternative set. We can see this in (7.34), where salient properties of the rheme alternative set include alternatives for each of *compete*, *fairly* and *their country*. Bringing this back to the discussion in section 7.1 about differences in the interpretation of relative givenness and kontrast; we would suggest that, in general, relative prominence structure within the theme or rheme phrase indicates the relative givenness of the elements of that phrase. The element that is not relatively given in relation to the phrase is the kontrast (mediated by prominence structure constraints). However, increased acoustic and structural prominence can induce a reading of salient alternative sets, or *restricted* kontrast, for other elements in the phrase that are particularly emphasised.

In all of the examples so far, the theme/rheme relationship has been indicated by ordering, i.e. theme comes before rheme. And so the last example shows the prosodic expression of rheme-theme ordering. As we saw in the experiment in section 6.4, theme-rheme order is partly cued by its frequency, since roughly two thirds of the information pairs there were theme-rheme. That is, rheme-theme is much less likely and therefore less expected. However, there were some examples in our dialogue, and we see that once theme/rheme status has been established from the prosody, the derivation of meanings works in exactly the same way. (7.35) carries on from the extract in (7.15):

(7.35) B: Germany's supposed to be doing something similar to that right now... they have jobs out on bulletin boards, so people know

B: **what is OPEN for an apprentices in different FIELDS**

We can see in Figure 7.13 that the nuclear accent on *open* has much higher pitch and intensity than the nuclear accents on *apprentices* and *fields* (sound file *apprentice*). In fact, the pitch and intensity over the whole first phrase is much higher than in the subsequent phrases. This marks the second two phrases as subordinate to the first; therefore the first is the rheme, and the last two the thematic, as in (7.36).

(7.36) ((what is OPEN_N)$_\rho$
((for an APPRENTICES_N) (in different FIELDS_N))$_\theta$ =>

(7.37) [[[$_S$ what is [open]$_F$ [for an apprentices in different fields]$_F$]]]^o
= { open(what_jobs,apprentices_in_different_fields) }

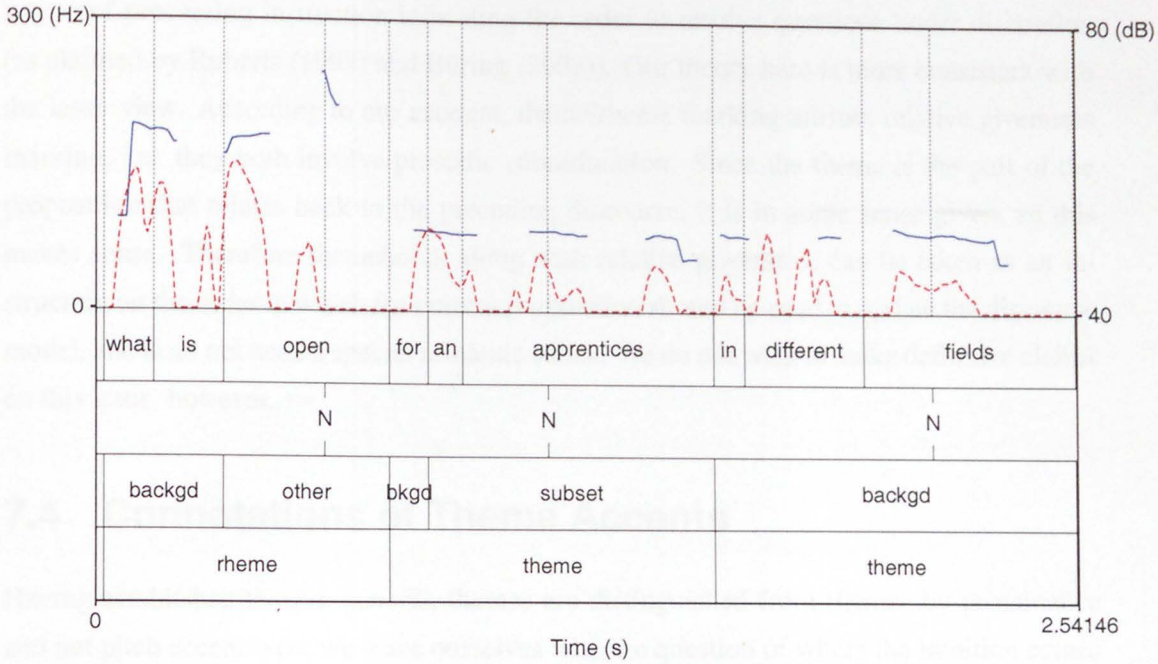


Figure 7.13: $f0$ trace (blue line) and intensity curve (dashed red line) for (7.35), along with the word transcript, accent type, kontrast status and theme/rheme annotation. Note theme/rheme annotation is per prosodic phrase.

(7.38) [[[what is open [for an apprentices in different fields]$_F$]]]$^{f\theta}$
= { open(what_jobs,x) | $x \in E$ }, where E is the domain of types of potential employees

(7.39) [[[what is [open]$_F$ for an apprentices in different fields]]]$^{f\rho}$
= { P(what_jobs,apprentices_in_different_fields) | $P \in E$ }, where E is the domain of states of availability

Once this marking has been established, the derivation works in exactly the same way, with an ordinary semantic value as in (7.37), and theme and rheme alternative sets in (7.38) and (7.39).

Overall, these examples seem persuasive support for our claim that theme/rheme status is indicated by relative prominence within metrical prosodic structure; and that this needs to be carefully separated from the marking of kontrast, particularly the invocation of a *restricted* kontrast interpretation, which is shown by increased prominence generally. On a final note, this evidence brings us back to the question raised in section 2.2.3 as to whether

the theme/rheme distinction is semantic (as claimed by Steedman 2000), or pragmatic, e.g. a type of processing instruction indicating the order to resolve questions under discussion (as claimed by Roberts (1998) and Büring (2003)). Our theory here is more consistent with the latter view. According to our account, theme/rheme marking mirrors relative givenness marking, i.e. they both involve prosodic subordination. Since the theme is the part of the proposition that relates back to the preceding discourse, it is in some sense given, so this makes sense. Therefore themehood, along with relative givenness, can be taken as an instruction on the order in which the current proposition should be used to update the discourse model, and does not need a special semantic status. We do not wish to make definitive claims on this issue, however.

7.4 Connotations of Theme Accents

Having established that, in general, themes are distinguished from rhemes by prominence and not pitch accent type; we leave ourselves with the question of where the intuition comes from that there is a certain type of accent, variously labelled 'scooped' or L+H*, that is only appropriate on themes. We still do not seem to be in a position to give a definitive answer to this question. However, on the basis of the evidence accumulated so far in this thesis, and in particular the discussion in section 3.3.2 on how meaning is conveyed by intonational tune, we present a speculative explanation, with reference to some examples from our dialogue.

In section 3.3.2, we argued that intonational 'morphemes' should be thought of as configurations of phonetic features operating at different levels of prosodic structure. From our results so far, we can refine the configuration in which these 'theme accents' (as we will call them here for the sake of clarity) occur. As we saw in the examples in the last section, they are not necessary to convey theme status in general. Further, as we saw from the results of the perception experiment in Chapter 4, they are not necessary for the perception of theme status. However, as we saw in the production experiment in Chapter 4, and again in the corpus study at the end of the last chapter, there are small, though fairly unstable, effects on the depth and location of the Low at the start of the accent, and the location of the Peak. These effects only occur on nuclear accents, and, we would like to claim, only when the accent is also emphasised to make its alternative set particularly salient; as they are because of the experimental set-up in Chapter 4, and, as we shall see, in all the relevant examples below. Therefore, we are looking at effects on accent shape only for emphasised accents in nuclear position.

From an analysis of our results so far, we can also say that the distinction between theme and rheme accents in these environments is not categorical. Firstly, as we saw in Chapter 4

in our own and other reported experiments, there is no robust difference in accent shape between theme and rheme accents (or L+H* and H*) over both production and perception. Secondly, the unstable but persistent differences we do find in production are not consistent with a 'categorical' shift. As we discussed in section 3.1.3.2, attested categorical shifts in Low or Peak location usually involve whole syllable differences, perhaps exploiting non-linearities in the physical system, as suggested in section 3.3.2. All the effects we found involve much more subtle differences within the stressed syllable.

Therefore the effects on emphasised nuclear theme accents which we are investigating here are variations on the realisation of a basic H* accent. In section 3.3.2 we discussed how pitch accent types can be thought of like phonemes, which vary considerably in their realisation because of the effects of context. However, unlike with phonemes, these variations can be meaningful at the same level of interpretation, i.e. discourse semantics. Further, we suggested that these variations may be directly meaningful, i.e. they are ideophonic, perhaps stemming from underlying 'biological codes' of intonation meaning. Here we will see if we can use this idea to explain our subtle 'theme accent' effects.

Figure 7.15 shows a selection of particularly emphasised theme and rheme nuclear accents taken from the extract in (7.40), as well as one (*defence*), taken from (7.5) to make an even set (sound files *defenceT*, *reallyT*, *goingT*, *enjoyT*, *dropoutR*, *enjoyR*, *likeR* and *lifeR*). All are said by B. (7.40) carries on from the extract in (7.1) where the speakers are discussing the benefits of educating kids through college, as opposed to apprenticeships:

(7.40) B: Well, it seems to me that kids that get out of high school, that parents have gone to college, and college here and college there
that are **REALLY** not interested in **GOING** to college, and forced into it, usually are your **DROP OUTS**
where if they're said 'hey', it's just as advisable to go into something you **ENJOY**, and you **LIKE**
because you can get just as far being a journeyman carpenter, or electrician, or a plumber... make as much money
and if they **ENJOY** it more, they make a happier **LIFE** for themselves

(7.41) - (7.44) show the information structure marking derived from the prosody in the same way as in the last section. From this we can see that *defence*, *really*, *going* and *enjoy it* are marked by thematic nuclear accents, and *dropouts*, *enjoy*, *like* and *life* by rhematic nuclear accents:

(7.41) ((as far as my DEFENCE$_N$) (budget))$_\theta$

(7.42) ((that are REALLY$_N$) (NOT$_N$) (interested in GOING$_N$ to college)
(and forced INTO$_N$ it))$_\theta$ (usually are your DROPOUTS$_N$)$_\rho$

(7.43) ((it's JUST$_N$ as) (ADVISABLE$_N$))$_\theta$
((to go into something you ENJOY$_N$) (and you LIKE$_N$))$_\rho$

(7.44) (and if they ENJOY$_N$ it more)$_\theta$
((they make a) (happier LIFE$_N$ for themselves))$_\rho$

In Figure 7.15, we can see immediately that theme accents do not form a coherent set that is distinct from rheme accents, consistent with our general claim. Obviously, since this is corpus data, we could not control the segmental content, so unvoiced regions do not have pitch tracks, etc. However, it is evident that there is much variation within both groups. Looking at the theme accents first, we can see that there is a high, definite peak on *defence* and *enjoy*, while the accents on *really* and *going to* are much flatter. Among the rheme accents, the peaks on *dropouts* and *like* are definitely earlier relative to the stressed syllable. However, this does not seem to hold for the samples in which we can make the most direct comparison. *enjoy* appears as both a theme and a rheme. In the rheme token, the peak is slightly earlier, relative to the stressed vowel, than in the theme; but the stressed vowel itself is longer, so it is difficult to tell if this is really an alignment effect. In its context, thematic *enjoy* is paired with *life*; and we can see that, between these two, alignment of the peak is identical, i.e. at the end of the stressed vowel. These two accents are also much higher than the rest of the set (pitch range shown is 0-380Hz, not 0-300Hz). In all cases where it can be clearly identified, the Low seems to be aligned just after the beginning of the stressed vowel.

In order to explain these effects, we are going to appeal to two of the 'modifications' of pitch accents suggested by Gussenhoven (1984), which we set out in section 2.3.1. Under his scheme, these modifications can be made to any basic pitch accent, adding nuances to its meaning. He claimed that increased *range* added greater 'insistence' to the speaker's meaning; and that *peak delay* added an implication that the element was 'non-routine' or 'especially significant'. Here we see that, among accents which are already perceived as emphasised, the ones that have particularly large pitch ranges do have an implication of 'insistence'. When B refers to the *defence* budget, she wants to insist that her view on this is different to that she has just expressed on the *prison systems*, i.e. her view the *defence budget* should not be cut further is very separate and final. Similarly, the final *enjoy* and *like* tokens come at the end of her long argument about why kids shouldn't be forced to go to college, an argument A doesn't appear to agree with. Further, among the tokens which do have comparatively late peaks, i.e. the two *enjoy*s and *life*, there is a definite implication of non-routineness or especial significance. B wants to emphasise that these kids would

enjoy their experience, as opposed to most kids who are either dropouts or miserably suffer through college. Under our current formulation, we can state this 'non-routineness' in terms of implying that it is especially significant that it is this member of the alternative set, and not any other, that is applicable, i.e. *enjoy* as opposed to *be miserable* and *better life* as opposed to *unemployment*. The tokens without late peaks do not have this especial significance. *really* is a strengthening modifier, without any salient alternatives; *going* and *dropouts* are given in the context, and *like* is a repeat of *enjoy*.

Applying this to our original question, we can suggest that it is not that emphasised kontrastive themes inherently have later peaks that kontrastive rhemes. Rather, kontrastive themes are more likely to have this 'especial significance' implication on their alternative set than kontrastive rhemes, signalled by peak delay. This seems plausible, since kontrastive themes, by definition, relate back to the preceding context. If a speaker wants to highlight a referent that is already established, it is likely that it is because it is especially significant that it is that referent, and not its alternatives, that is used in the context. Going back to our discussion at the beginning of this section, we do not expect this 'peak delay' to be categorical. Rather, the greater the peak delay, the greater the likelihood of this 'especial significance' reading, mediated by the plausibility of such a reading in the context. In this way, peak delay is gradient and 'directly meaningful'. Also, it is probable that 'peak delay' is only the most easily observable correlate of the phonetic variations that lead to the perception that these accents sound 'scooped'.

On the other hand, there may be a categorical effect marking rhemes, as opposed to themes. In most of the literature, and in this thesis, discussion centres on the prosodic signals which separate thematic or 'contrastive' accents from rhematic or 'ordinary' focus accents. There has therefore not been as much attention given to positive signals of rhemehood. We saw in the production experiments in Chapter 4 that nearly all rhemes are followed by a drop in $f0$ and a flat or falling boundary, while the continuation from a theme accent is much more variable. In the examples in this chapter, nearly all rheme phrases are followed by a definite drop in pitch; while in theme phrases, this drop is not so severe, or the boundary is rising (e.g. (7.6) and (7.30)). As we suggested in section 4.2.4.2, the most likely explanation for this is the marking of the nuclear accent at the higher phrasal level, since nuclear accents are often followed by a drop in $f0$ (see section 3.1.2). This marking may also follow from generally held views on the 'meanings' of rising and falling boundaries, i.e. that rising boundaries are 'forward-looking' (i.e. to fill an open proposition), and falling boundaries are 'final' (i.e. marking a complete proposition). However, this would be more difficult to reconcile with evidence that the *depth* of the fall is significant, i.e. a greater fall, not a fall per se, marks rhemehood.

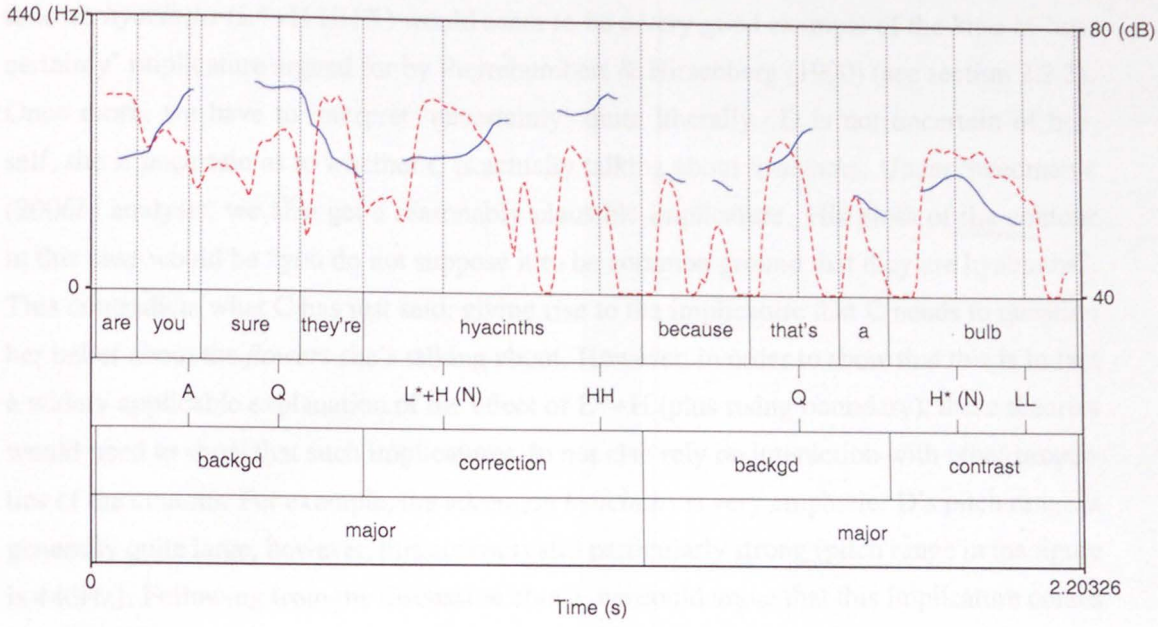


Figure 7.14: $f0$ trace (blue line) and intensity curve (dashed red line) for (7.45), along with the word transcript, accent type and ToBI transcription, kontrast status and phrase type annotation.

Finally, we may ask how this analysis affects existing theories about how other pitch accent types signal 'meaning', as discussed in section 2.3. As we discussed in section 3.1.3.2, the categorical basis on which other ToBI pitch accent types are defined (at least L*, L*+H and H*) still stands. However, any future investigation of their meaning will need to control carefully for the information structure, kontrastive interpretation, as well as the context in which they appear, before such theories can be verified. We will briefly discuss what we mean by this in relation to one last example, which looks at the 'negative polarity' meaning associated with L*+H, discussed in section 2.3.3. This example, from a different conversation, is repeated from (5.10), as, in the author's opinion, there were not any good examples of this sort of implicature arising from L*+H in the conversation used in the rest of this chapter:

(7.45) (C)... it was a hyacinth have you ever seen those? Oh they are pretty in the Spring but the leaves I do not like them...
(D) now **are you sure they're HYACINTHS because that is a BULB**

Figure 7.14 shows the acoustic representation, along with the kontrast and phrase annotation, and a possible ToBI transcription of the nuclear accent and following boundary tone

in each phrase (sound file *hyacinths*). On the face of it the accent and following boundary tone on *hyacinths* (L*+H HH%) would seem to be a very good example of the kind of 'uncertainty' implicature argued for by Pierrehumbert & Hirschberg (1990) (see section 2.3.3). Once more, we have to interpret 'uncertainty' quite liberally. D is not uncertain of herself, she is uncertain as to whether *C* is actually talking about *hyacinths*. Under Steedman's (2006*b*) analysis, we also get a reasonably plausible implicature. His gloss of the contour in this case would be "you do not suppose it to be common ground that they are hyacinths". This contradicts what C has just said, giving rise to the implicature that C needs to question her belief about the *flowers* she's talking about. However, in order to show that this is in fact a widely applicable explanation of the effect of L*+H (plus rising boundary), these theories would need to show that such implicatures do not also rely on interaction with other properties of the contour. For example, the accent on *hyacinths* is very emphatic. D's pitch range is generally quite large, however, this accent is still particularly strong (pitch range in the figure is 440Hz). Following from the discussion above, we could argue that this implicature comes as much from the connotations arising from emphasis, as from L*+H itself. In particular, we argued that peak delay on emphatic accents could signal that it was 'especially significant' that the speaker referred to this member of the alternative set, as opposed to any other. This analysis certainly seems to apply here, as D wants to question whether *hyacinths* is the right member of the alternative set of *flowers*. The 'negative' or 'uncertain' implicature could arise as much from the words themselves, i.e. *now are you sure*. As discussed in section 4.3, it is not clear how Steedman's analysis that these are 'isolated themes' would account for the H* accent on *you*. This particularly strong pre-nuclear accent could be argued to signal a kontrast on such a thematic, pronominal element. This would result in negative and positive polarity being signalled in the same theme phrase. Finally, it could be argued that D's final phrase, *that's a bulb*, has a similar status to *he's a good badminton player* (discussed in section 2.3.3). That is, it is being offered as negative evidence against an assertion of the other speaker. However, here the L*+H accent is not used, but (in the author's opinion at least) the implication that C should know this still holds, which is the reason since utterances are analysed as thematic. On the other hand, it is said with much lower pitch, consistent with our suggestion that the rhetorical relationship of Nucleus-Evidence may be signalled by prosodic subordination independently of theme/rheme status (cf. Mann & Thompson 1988). None of these arguments are meant to be conclusive; they are presented to illustrate the types of evidence such theories must account for if they are to show that the 'meanings' of these pitch accents and boundary tones really have as broad a coverage as is claimed.

Overall, the results in this chapter nicely confirm the predictions of our theory relating information structure and metrical prosodic structure in Chapter 3. We saw that relative givenness, i.e. givenness in relation to a proposition, is consistently signalled by relative prominence in the metrical structure, while there is no consistent effect of accenting. Once recursive phrasing structure is taken into account, theme and rheme scope is straight-forwardly determined by phrase boundaries, while syntactic projection theories fail when prosodic and syntactic phrasing differ. We saw some indication that increased prominence can affect the interpretation of kontrast within the theme or rheme phrase, however this is mediated by the semantic and prosodic context in the way predicted. Theme/rheme status was again shown to be signalled by relative prominence across prosodic phrases; and *restricted* kontrast by increased prominence on nuclear accents. Finally, we suggested that the distinctive accent shape often claimed for theme accents in fact comes from the 'especial significance' implication of peak delay. That is, in these cases, the speaker is not trying to mark the fact the accent is thematic, but that it is especially significant that they are referring to this theme, and not others in its alternative set. Taken together with the more broad-ranging findings in the last chapter, and the experimental results in Chapter 4, this analysis provides persuasive evidence for the general theory being advanced here.

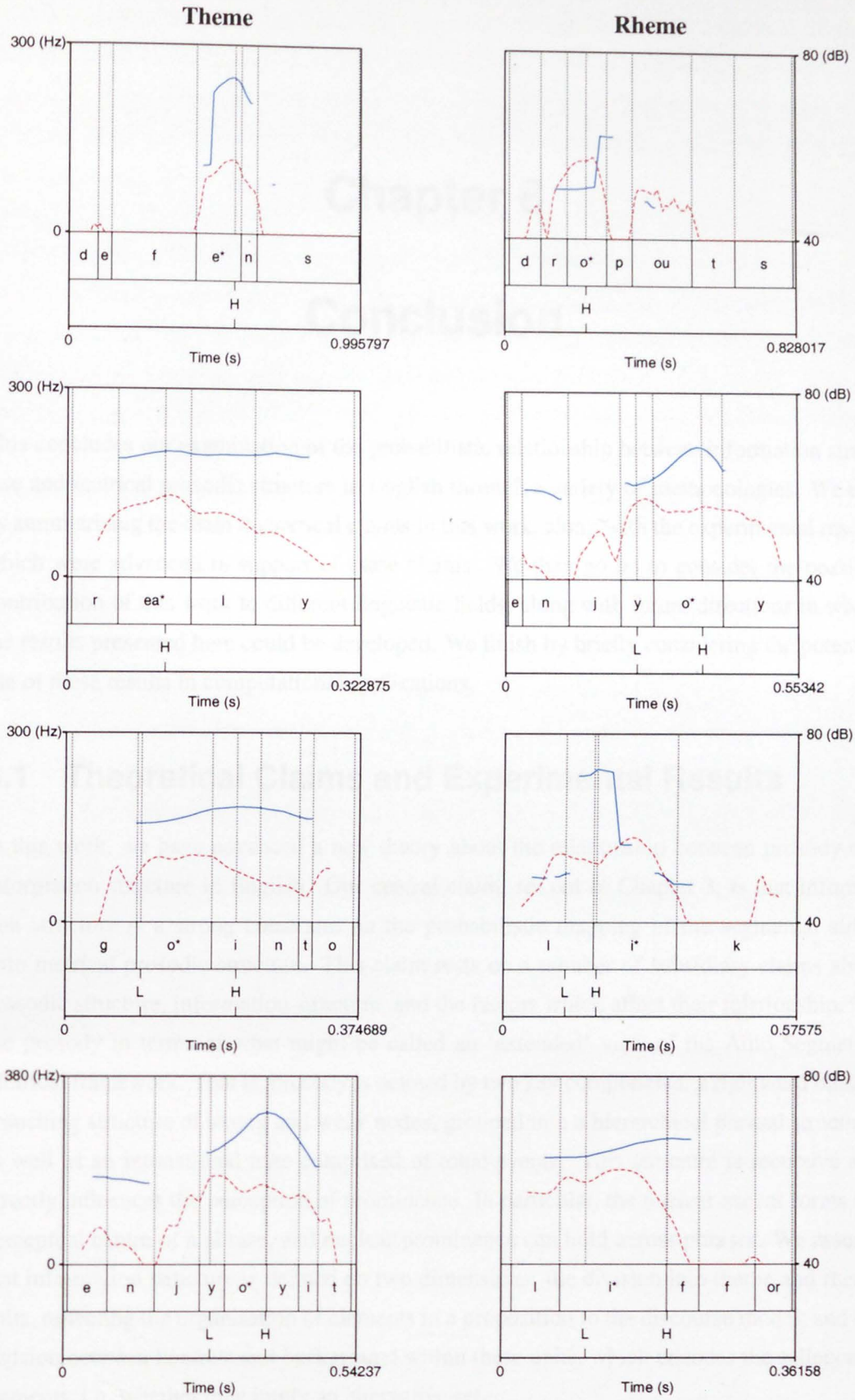


Figure 7.15: $f0$ (blue line) and intensity (dashed red line) for theme and rheme nuclear accents in (7.5) (*defence*) and (7.40), with the phones (* = stressed) and location of the start of the accent rise (L) and the accent peak (H).

Chapter 8

Conclusion

This concludes our examination of the probabilistic relationship between information structure and metrical prosodic structure in English through a variety of methodologies. We end by summarising the main theoretical claims in this work, along with the experimental results which were advanced in support of these claims. We then go on to consider the possible contribution of this work to different linguistic fields, along with future directions in which the results presented here could be developed. We finish by briefly considering the potential use of these results in computational applications.

8.1 Theoretical Claims and Experimental Results

In this work, we have advanced a new theory about the relationship between prosody and information structure in English. Our central claim, set out in Chapter 3, is that information structure is a strong constraint on the probabilistic mapping of the segmental string onto metrical prosodic structure. This claim rests on a number of subsidiary claims about prosodic structure, information structure, and the factors which affect their relationship. We see prosody in terms of what might be called an 'extended' view of the Auto-Segmental Metrical framework. That is, prosody is defined by two key components: a rightward binary-branching structure of strong and weak nodes, grouped into a hierarchical phrasal structure; as well as an intonational tune comprised of tonal events. This structure is recursive and directly influences the perception of prominence. In particular, the nuclear accent forms the perceptual centre of a phrase, and nuclear prominence can hold across phrases. We assume that information structure is defined on two dimensions: the division into theme and rheme units, reflecting the organisation of elements in a proposition to the discourse model; and the division between kontrast and background within these units, which encodes the salience of elements, i.e. whether they imply an alternative set.

The mapping of words onto prosodic structure is strongly constrained by these information structural properties. That is, kontrastive elements 'try' to align with nuclear accents; and the scope of focus (i.e. theme/rheme units) is determined by the scope of the prominence that the kontrast is aligned with. Theme/rheme status is signalled by relative prominence. If there is a kontrast within both the theme and the rheme, then the theme nuclear accent will be less relatively prominent than the rheme nuclear accent. However, this mapping is also affected by other constraints on the realisation of prosodic structure, including the word class of the elements involved, syntax structure, referent accessibility, and the signalling of higher level 'meanings' such as emphasis. There are also constraints inherent in prosodic structure itself, including production pressures and rhythmical requirements. Therefore we argued that this mapping is probabilistic. Pre- and post-nuclear accents can signal kontrast subject to their other properties in context. The corollary of our claim is that intonational tune is much less important to signalling information structure than had previously been thought. In Chapters 2 and 3, we laid out evidence from the literature supporting all of these aspects of our theory. In particular, we showed that our approach is better able to account for a range of cases that were problematic for previous theories which presumed that kontrast/focus is marked by pitch accenting per se, that the scope of focus is determined syntactically, and that the distinction between theme and rheme kontrast is signalled by tonal event type.

In the later part of the thesis, we tested whether the predictions of our theory would hold over a broad range of language, i.e. a corpus of unrestricted spontaneous speech. This involved the development of the Switchboard corpus, which had already been annotated for POS/syntax, disfluencies, dialog acts, information status and phones/syllables. Using the NXT system, we added substantial new layers of kontrast and prosodic features, including nuclear accents, as described in Chapter 5. These were reasonably successful. In Chapter 6, we then used a series of multiple logistic regression and CART models to show that the distributional properties of the corpus were consistent with our claims. In Chapter 7, we demonstrated more fine-grained aspects of our theory using examples from the corpus.

The first part of our claim to be tested was that kontrast aligns with nuclear prominence. Since we assume that the position of the nuclear accent within a phrase is strongly constrained by right-branching prominence, this implies that information units place a strong constraint on prosodic boundaries. That is, information units are marked by prosodic boundaries at some level of structure. These constraints interact with the other known constraints on prominence and phrasing. In Chapter 6, using phrase break prediction models, we showed that clause and constituent structure act as strong constraints on phrasing, along with positional and rhythmical factors, and to a less demonstrable extent kontrast and information status. Importantly, we showed accentual features had little effect on phrase break predic-

tion, consistent with phrasing regulating prominence, and not vice versa. In Chapter 7, we gave a range of examples from the corpus where focal structure could be neatly determined by phrasing, but was in conflict with syntactic focus projection rules. These findings formed the basis for the next series of accent prediction models in Chapter 6. We showed that plain and nuclear accents can best be distinguished by their phrasal properties, again confirming the right-branching bias. Crucially, nuclear accents can be reliably predicted by semantic/syntactic features, especially kontrast, while other accents cannot. Plain accents, however, are more likely to fall on syntactically 'strong' words, e.g. nouns or objects, while these features are not useful with nuclear accents. Further, plain accent prediction substantially improved when word-level acoustic features were used, while nuclear accent prediction did not. These findings are all consistent with the claim that nuclear accents are directly 'meaningful', i.e. kontrastive, while plain accents are usually not. In Chapter 7, we showed many examples where kontrast/relative givenness was signalled by nuclear accenting, with the appearance of other accents being largely unrelated to focal structure.

Our theory went further than this kontrast/nuclear prominence relationship, however. Since the mapping between the two is probabilistic, other accents can signal kontrast, and not all nuclear accents mark kontrast. The guiding principle is that a kontrastive interpretation is more likely if an element is more prominent, both structurally and acoustically, than expected given its properties. We further claimed the more prominent a kontrastive element, the more likely a *restricted* kontrast interpretation, given the context. Results from our kontrast prediction models in Chapter 6 were directly in line with these predictions. A kontrast is more likely if a word carries a nuclear accent. However, a kontrast is also more likely if the word is in a more prominent accent position (i.e. nuclear or accented) than would be expected from its syntactic/information status properties; and if it is more acoustically prominent than would be expected given these and its structural prominence. Chapter 7 showed a number of examples where a particularly strong pre-nuclear accent signalled kontrast, and where kontrast perception can arise from nuclear prominence perception over multiple phrases. We also showed how the perception of a restricted alternative set can arise from both increased prominence and context.

Inherent in our claims about structural prominence was the idea that the marking of acoustic prominence depends on structure. Post-nuclear elements have much lower pitch than nuclear elements. Pre-nuclear elements may be more acoustically prominent than nuclear accents, however the nuclear accent still forms the 'perceptual centre' of the phrase. Therefore, nuclear accents are more likely to show meaningful variation in pitch accent shape. Our findings on the acoustic properties of pre-, post- and nuclear accents in Chapter 6 were consistent with these claims. Post-nuclear accents have much lower pitch than

the other two, and nuclear accents are longer and have greater pitch range than other accents. However, as we saw before, these acoustic differences are not necessary for nuclear accent perception, and so may be manipulated to convey information structure. We showed that kontrastive nuclear accents have higher pitch, pitch range, duration and later peaks than backgrounded nuclear accents. Finally, in Chapter 7 we saw examples of nuclear accents which had much lower pitch than their pre-nuclear accents; and suggested this may be correlated with discourse givenness.

The final part of our story was the claim that kontrastive themes are distinguished from kontrastive rhemes by relative prominence, i.e. the rheme is nuclear at a higher phrasal level; rather than by tonal event type, e.g. L+H* versus H*. As we showed in Chapter 2, this question needs to be carefully distinguished from how 'contrastive' accents, i.e. *restricted* kontrast, are distinguished from other accents, which we had already established is correlated with increased prominence in general. In Chapter 4 we showed that when contrastive theme accents are compared with contrastive rheme accents, theme peaks are lower. In a production study, we showed a number of subtle accent shape distinctions between the two, including peak height. However, in a complementary perception experiment, the only factor listeners could reliably use to judge the acceptability of theme accents was peak height. In a follow-up production experiment, we directly compared paired theme and rheme accents, and showed the height distinction reflected our understanding of the signalling of nuclear prominence: theme peaks were only lower than rheme peaks in rheme-theme order, consistent with post-nuclear lowering. The smaller corpus study in Chapter 6 confirmed this result. In Chapters 3 and 7, we explored where the intuition that themes are marked by a distinct pitch accent could have come from. We suggested that it may arise from more subtle variations in pitch accent shape signalling affective connotations correlated with themehood. In particular, we suggested that peak delay might signal that it is 'especially significant' that it is that member of the theme alternative set, and not contextually available others, that is used.

8.2 Contribution and Future Directions

This work has shown, through a wide variety of methods, the importance of metrical prosodic prominence, as opposed to intonational tune, in signalling information structure. More generally, it contributes further evidence for the *stress-first* theory of phrasal stress, and the superiority of stress-first explanations of focus marking (Ladd 1996, Truckenbrodt 1995, Büring to appear, Wagner 2006). We have shown the centrality of nuclear accents to the meaning of utterances. It is hoped that this sort of evidence will lead to a re-evaluation of the status of 'accents' across linguistic work. Evidently, these cannot be considered to be

a uniform phenomenon. Our results show not only the distinctive properties of nuclear accents, but suggest that the perception of other accents in fact draws a somewhat arbitrary boundary between levels of relative prominence in a phrase. Further, there are meaningful relative prominence distinctions between nuclear accents. The results in regard to this part of our claim are essentially only from language production. In the future, it would be useful to carry out perception experiments along the lines of that in Chapter 4, to test whether kontrast and *restricted* kontrast are perceived through increased prominence as claimed, and to test directly whether relative prominence signals status in paired theme/rheme elements. In general, we submit that future work within discourse semantics looking at the importance of accent distribution and tonal pitch accent type in signalling both information structure and the type of illocutionary and affective connotations discussed here, needs to carefully control for the relative prominence of the 'accents' involved. This work also has implications for prosodic annotation of corpora. Ideally, it would be useful to capture the expression of relative prominence and recursive structure more directly. One way to begin could be to experimentally evaluate how many levels of prominence can be reliably distinguished by annotators (see also Dilley 2005). We will see in the next section that this also relevant to automatic pitch accent prediction.

Another central claim is that the relationship between prosodic structure and information structure is probabilistic. This was borne out by the results in Chapter 6 and shown anecdotally through examples in Chapter 7. It is hoped, as discussed in section 3.3.1, that this work will contribute to the growing body of research showing that human language processing can be conceived in terms of constraint-based probabilistic models. The statistical models built in this thesis were meant to provide a way of showing the different constraints on the realisation of prosody and kontrast. However, it would be interesting to develop the implementation of some of these ideas sketched in Figure 3.15 in terms of formal constraint-based models (e.g. Tabor et al. 1997) (see further in Jurafsky 2003). The probabilistic relationship we have argued for also suggests an explanation for some cross-linguistic differences in the signalling of, e.g. givenness, found in the literature. For instance, Ladd (1996, ch. 5) shows that in Romance languages, such as Italian, it is not usual to deaccent repeated mentions, so *I bought her WHISKY, but she doesn't like WHISKY* would sound perfectly acceptable. This could be quite straight-forwardly accommodated in terms of the relative strength of the constraints involved, i.e. making given items less relatively prominent, and preserving right-branching structure. It is at least starting point for future research to look in corpora to see if these differences are absolute, or tendencies in each language.

In this thesis, we have largely assumed the categorial status of the informational structural concepts which we were trying to model. The reasonable success of our prediction

models can be seen as indirect support for this assumption. However, as we discussed in section 2.2.2.2, there is another quite separate body of research relating the prominence of elements in a discourse to more gradient notions such as referent accessibility, informativity and/or predictability, with some even claiming these subsume the role of focus. As we saw in Chapter 6, the information status of referents as annotated in our corpus did not seem to be a very useful feature for accent status prediction. Moreover, it is difficult to see how the depth and clarity of the discourse semantic interpretation obtained from information structure, particularly the implications of alternative sets and focus scope seen in the examples in Chapter 7, could be captured through such gradient notions. However, the experimental method used in Chapter 6 could be extended to build relative prominence prediction models comparing traditional information structural features against standard predictability and informativity measures used in the computational field, such as unigram and bigram frequency and TF*IDF. On the other hand, our results do provide reasonably direct support for the structural nature of prosody. In particular, we saw in the kontrast prediction models in Chapter 6 that straight acoustic features did not perform better than accent status, showing the expressive power of different degrees of prominence goes beyond their acoustic properties.

At several points in the thesis, we have raised the status of the relationship between information units, prosodic phrasing and syntax structure. We have claimed that information units strongly constrain the placement of phrase boundaries, i.e. information units are marked by phrase boundaries at some level of recursive phrasing structure. As we noted, it was not possible to confirm this using our prediction models in Chapter 6. However, we did find that constituent and clause boundaries very strongly predict phrase boundaries. On the other hand, phrasing regulates the perception of nuclear prominence, and nuclear prominence is strongly correlated with kontrast, which in turn delimits information units. This is suggestive of an least two effects. The first, as we have alluded to, is that prosody itself constrains syntactic parsing, as claimed by Steedman (2001) in his Combinatory Categorial Grammar. It would be interesting to test our phrase prediction model using syntactic features derived from a corpus annotated with this grammar, to see if his, and our, claims about the relationship between information structure and phrasing hold up. Secondly, the circular relationship described above suggests that it is syntactic ordering which is being manipulated so as to place kontrastive elements in the default nuclear position within prosodic phrases. Using our current corpus, it would certainly be possible in the future to look at the relationship between known syntactic alternations, e.g. *passive/active* or *ditranstive objects*, information structural status, and prosodic phrasing and prominence.

As we have said, one consequence of our theory is that intonational tune is much less important than previously thought to signalling information structure. However, we did not look

closely at other proposed informational structural properties, e.g. *mutual belief/polarity in relation to the common ground* and *speaker/hearer orientation/supposition*, which are meant to lead to the illocutionary and affective connotations claimed in section 2.3.1. As we discussed in section 3.3.2, the prosodic picture in relation to the accents claimed to signal these meanings is more complex than most of these theories presume. Along with the 'meaning' of a phonological low accent, for example, low pitch may inherently convey affective 'meanings', as may subtle variations in accent shape. Further, the meaning of such accents may be partly compositional, and partly derived from the meaning of semi-lexicalised intonational tunes. Lastly, as we discussed at the end of Chapter 2, intonational 'meanings' tend to be complex themselves, i.e. they can arise from configurations of signals at different levels of linguistic structure, rather than directly from a particular pitch accent. With regard to all of these considerations, we would suggest that research into the prosodic signalling of these 'higher-level' meanings should ideally take an approach similar to the one here, i.e. use annotated corpora so that other diverse features can be controlled for. However, it is inherently difficult to reliably annotate such features, particularly affective connotations (see Scherer & Banziger 2004), so this may not be feasible. It may be more profitable to look at illocutionary force, e.g. at dialogue acts in a corpus which displays a greater variety of these than Switchboard. The other possible route is further experiments along the lines of Scherer et al. (1984) and Ladd et al. (1985), which directly test the interaction of different levels of prosodic signals.

Finally, a note on the general methodology and approach taken in this thesis. As discussed at the end of Chapter 2, we believe it is important to test the generalisability of theories developed using introspective examples to a wider range of language. We have seen in our work how diverse linguistic factors can interact in natural language, so that the apparently direct relationships between information structural properties and prosodic structures break down. More generally, we noted in discussion in Chapter 5 that it is not always straight-forward to apply theoretical concepts such as kontrast to unrestricted language. While the models reported in Chapter 6 show that our approach was successful to capturing the prosodic marking of kontrast as annotated; the level of annotator agreement in Chapter 5 suggests that more qualitative analysis of the information structure of problematic examples taken from real discourse (along the lines of Chapter 7) would be useful to refine the notion of kontrast, and importantly focus scope, to be able to more reliably capture these cases. Further, we would suggest that inherent in the very nature of our model is the idea that it does not make sense to study prosody, at least the nature of prosodic phonology, divorced from meaning. Take for example the issue of tonal target alignment with syllables. Evidence of a 'categorical' break in the perception of such alignment, i.e. from one syllable to the next,

is evidence that such alignment tends to be used and interpreted discontinuously. However, it does not imply at what level this discontinuity is perceived, and therefore the phonological status of the shift.

8.3 Applications

We conclude with a brief look at the potential applications of this work beyond usual linguistic concerns. As we noted at the end of Chapter 6, our results there have direct implications for pitch accent prediction systems, i.e. as we discussed, it is probably not useful for natural language applications to work on the prediction of pitch accents per se. Rather, models need to be built which can, at least, distinguish nuclear from plain accents, and more attention needs to be paid to phrase break prediction below the sentence level.

It is also hoped that these results may be useful to improve prosody in speech synthesis systems. The current state-of-the-art is unit selection synthesis, i.e. in producing a particular utterance, units of varying length are selected from a large database of speech annotated for different relevant features. The general idea is for there to be as little manipulation of the original speech signal in generating the synthesised output as possible. In carefully controlled domains, or in contexts involving 'neutral' prosody, the results are undoubtedly better than older methods which used direct manipulation of the speech signal; and are often indistinguishable from natural speech (e.g. see Clark & King 2006). However, as Clark & King (2006) note, such systems break down, i.e. the output sounds very unnatural, when the synthesised sentences are outside the domain of those in the database, particularly when the meaning of the utterance entails non-'default' prosody. The solution is to specify, along with features such as the identity of the phone, and its phrasal position, the prosodic features of each unit; so that prosody can be generated at the same time as unit selection (e.g. Clark & King 2006). Unfortunately, as these authors note, the number of prosodic features needs to be small or the size of the resulting database would be unmanageable.

The results in this thesis would suggest that rather than specifying pitch accent type, the concentration should be on levels of prosodic prominence (though boundary tones would probably still need to be specified). Most of the meaningful distinctions discussed in the earlier chapters could be captured reasonably well in terms of a three-way classification into weakly accented, nuclear accent and emphatic accent. Further, rather than the rather rough features currently being used to predict 'meaningful' accents within databases used for unit selection systems (e.g. see Strom, Clark & King 2006); our kontrast prediction models, developed in Chapter 6, offer at least a starting point for predicting the occurrence of nuclear accents from semantic/syntactic features. For natural language generation systems,

our results are more immediately applicable. The identification of kontrast/background and theme/rheme in a sentence should be relatively straight-forward, as the system can keep track of which parts of a proposition are 'new' and which relate to the preceding discourse; as well as which referents have alternative sets (cf. Prevost 1995, Baker, Clark & White 2004). These features can then be used to predict the marking of weak, nuclear and emphasised accents in the database.

We would further hope that the results of any such implementation could be fed back into the investigation of linguistic questions raised in this thesis. At the very least, any evaluation of the acceptability of prosody produced using such a system could also be used to test the claims we have made about the semantic import of different prosodic patterns. In general, it would be good to see results from linguistic work informing computational work, and vice versa, more in the future.

Finally, it is hoped that the integrated Switchboard corpus in NXT, with its many layers of assorted linguistic annotation, will prove useful to others in the investigation not only of the relationship between prosody and information structure, but other diverse questions about language not envisaged by its current developers. This may potentially be through the addition of more layers of annotation within the NXT framework. To that end, we are looking toward its public release in the near future.

Appendix A

Stimuli for Experiment 1

A.1 Block 1

1. Q: Don't you have to be very fit to climb Ben Nevis?
   A: No, Ben Nevis is an easy climb.

2. Q: Isn't that book by Alan Lowry?
   A: It's by Anna Lowry, not by Alan Lowry.

3. Q: What method did the psychiatrist use?
   A: He has tried a course of hypnosis.

4. Q: That's Jane Vanderberg, isn't it?
   A: It's not Jane Vanderberg, it's Jane Mulder.

5. Q: That's money laundering you're suggesting!
   A: It's just a financial solution to the problem, not money laundering.

6. Q: Do you think 'The Matrix' was an arthouse or an indie film?
   A: I don't know, I haven't seen 'The Matrix'.

7. Q: Where is her place again? In Longmore?
   A: It isn't in Longmore, it's in London.

8. Q: Which is the coldest month of the year?
   A: Probably either January or February.

9. Q: That piece comes from Norma Munroe, doesn't it?
   A: It's not from Norma Munroe, it's from Norman Munroe.

10. Q: Who was in charge of planning the scheme?
    A: Jeremy McConville headed the team.

11. Q: Didn't you tell me that she had some monkeys?
    A: I didn't know she had some monkeys, I knew she had some wombats.

12. Q: She's from Havana, isn't she?
    A: She's from Malaya, not from Havana.

13. Q: What are the common symptoms of chicken pox?
    A: Red dots on the skin are common signs.

14. Q: That guy's Henry Lambert, I think.
    A: That's Henry Lombard, not Henry Lambert.

A.2 Block 2

1. Q: I'm just suggesting a financial solution to the problem...
   A: It isn't a financial solution to the problem, it's money laundering.

2. Q: Henri plays for Arsenal not Leeds, doesn't he?
   A: Yeah, he plays for Arsenal.

3. Q: Where is her place again? In London?
   A: It's in Longmore, not in London.

4. Q: Where does organic food come from?
   A: It comes from Greenock.

5. Q: She's from Malaya, isn't she?
   A: She isn't from Malaya, she's from Havana.

6. Q: Have you seen Jim lately?
   A: No, Jim's doesn't live in Edinburgh anymore.

7. Q: That guy's Henry Lombard, I think.
   A: That isn't Henry Lombard, it's Henry Lambert.

8. Q: What do you think is the best mountain to climb in Scotland?
   A: Ben Nevis is one of the best and most managable.

9. Q: Isn't that book by Anna Lowry?
   A: It's not by Anna Lowry, it's by Alan Lowry.

10. Q: Who had on their new high heels?
    A: Kate was wearing her new Jimmy Choos.

11. Q: That's Jane Mulder, isn't it?
    A: It's Jane Vanderberg, not Jane Mulder.

12. Q: Do you think the weather's worst in January?
    A: No, I think February can be more bitter.

13. Q: That piece comes from Norman Munroe, doesn't it?
    A: It comes from Norma Munroe, not from Norman Munroe.

14. Q: Didn't you tell me that she had some wombats?
    A: I thought she had some monkeys, not some wombats.

A.3 Block 3

1. Q: That guy's Henry Lambert, I think.
   A: That isn't Henry Lambert, it's Henry Lombard.

2. Q: Where is Jim from originally?
   A: Jim's from Edinburgh.

3. Q: She's from Havana, isn't she?
   A: She isn't from Havana, she's from Malaya.

4. Q: Why do adults have to be wary of red dots on the skin?
   A: Chicken pox as an adult can be deadly.

5. Q: Where is her place again? In Longmore?
   A: It's in London, not in Longmore.

6. Q: Do you think it's worth trying hypnosis?
   A: I don't think hypnosis is worthwhile.

7. Q: That's money laundering you're suggesting!
   A: It's not money laundering, it's just a financial solution to the problem.

8. Q: Who does Henri play for?
   A: He plays for Arsenal.

9. Q: Didn't you tell me that she had some monkeys?
   A: I knew she had some wombats, I didn't know she had some monkeys.

10. Q: That piece comes from Norma Munroe, doesn't it?
    A: It comes from Norman Munroe, not from Norma Munroe.

11. Q: What days are the classes run?
    A: The classes are run on Mondays, Wednesdays and Fridays.

12. Q: That's Jane Vanderberg, isn't it?
    A: It's Jane Mulder, not Jane Vanderberg.

13. Q: What was Jeremy's role in the process?
    A: Jeremy McConville headed the management team.

14. Q: Isn't that book by Alan Lowry?
    A: It's not by Alan Lowry, it's by Anna Lowry.

A.4 Block 4

1. Q: Didn't you tell me that she had some wombats?
   A: I didn't know she had some wombats, I thought she had some monkeys.

2. Q: What time is the movie, 8 o'clock? A: No, the movie starts at 9 o'clock tonight.

3. Q: That piece comes from Norman Munroe, doesn't it?
   A: It isn't from Norman Munroe, it comes from Norma Munroe.

4. Q: Why did Kate look so sad last night?
   A: She broke her new Jimmy Choos.

5. Q: That's Jane Mulder, isn't it?
   A: It's not Jane Mulder, it's Jane Vanderberg.

6. Q: Why is Greenock popular with hippies?
   A: Organic food comes from Greenock.

7. Q: Isn't that book by Anna Lowry?
   A: It's by Alan Lowry, not by Anna Lowry.

8. Q: What's your favourite film of the past few years?
   A: Definitely 'The Matrix'.

9. Q: That guy's Henry Lombard, I think.
   A: That's Henry Lambert, not Henry Lombard.

10. Q: She's from Malaya, isn't she?
    A: She's from Havana, not from Malaya.

11. Q: What days does Barry have off work?
    A: Barry doesn't work on Mondays, Wednesdays and Fridays.

12. Q: Where is her place again? In London?
    A: It isn't in London, it's in Longmore.

13. Q: Do you agree with his suggestion to use hypnosis?
    A: No, I think meditation is a better treatment than hypnosis.

14. Q: I'm just suggesting a financial solution to the problem...
    A: It's money laundering, not a financial solution to the problem.

Appendix B

Stimuli for Experiment 4

B.1 Block 1

1. A: So, are you mailing me the manuscripts?
   B: No, I'm handing you the manuscripts. I'm mailing you the magazines.

2. A: You're going to see Amanda on Monday, right?
   B: No, I'm seeing Amanda tomorrow, I'll see Norma on Monday.

3. A: Where are these limes from? Gautemala?
   B: No, the limes are from Australia. The mangoes are from Gautemala.

4. A: Are you interested in the Olympus camera with 90 mega pixels?
   B: No, I want either the Minolta with 90 pixels or the Olympus with 120 pixels.

5. A: OK, you want to catch the 20.30 train to London, then?
   B: No, I want to catch the 20.30 to Manchester and then the 21.30 to London.

6. A: Cool, you're going to Vienna in January?
   B: No, I'm off to Barcelona in January. I'm going to Vienna in November.

7. A: OK, you want to catch the 19.30 train to Manchester, then?
   B: No, I want to catch the 19.30 to London and then the 20.30 to Manchester.

8. A: Are you interested in the Olympus camera with 120 mega pixels?
   B: No, I want either the Olympus with 90 pixels or the Minolta with 120 pixels.

9. A: So, are you handing me the manuscripts?
   B: No, I'm handing you the magazines. I'm mailing you the manuscripts.

10. A: Where are these limes from? Australia?
    B: No, the limes are from Gautemala. The mangoes are from Australia.

11. A: Cool, you're going to Barcelona in January?
    B: No, I'm off to Vienna in January. I'm going to Barcelona in November.

12. A: You're going to see Norma tomorrow, right?
    B: No, I'll see Norma on Monday, I'm seeing Amanda tomorrow.

B.2 Block 2

1. A: You're going to see Amanda on Monday, right?
   B: No, I'll see Norma on Monday, I'm seeing Amanda tomorrow.

2. A: Are you interested in the Olympus camera with 90 mega pixels?
   B: No, I want either the Olympus with 120 pixels or the Minolta with 90 pixels.

3. A: So, are you handing me the manuscripts?
   B: No, I'm mailing you the manuscripts. I'm handing you the magazines.

4. A: Where are these limes from? Australia?
   B: No, the mangoes are from Australia. The limes are from Gautemala.

5. A: OK, you want to catch the 21.30 train to London, then?
   B: No, I want to catch the 20.30 to London and then the 21.30 to Manchester.

6. A: Cool, you're going to Barcelona in January?
   B: No, I'm going to Barcelona in November. I'm off to Vienna in January.

7. A: So, are you mailing me the manuscripts?
   B: No, I'm mailing you the magazines. I'm handing you the manuscripts.

8. A: OK, you want to catch the 20.30 train to London, then?
   B: No, I want to catch the 19.30 to London and then the 20.30 to Manchester.

9. A: You're going to see Norma tomorrow, right?
   B: No, I'm seeing Amanda tomorrow, I'll see Norma on Monday.

10. A: Where are these limes from? Gautemala?
    B: No, the mangoes are from Gautemala. The limes are from Australia.

11. A: Are you interested in the Minolta camera with 120 mega pixels?
    B: No, I want either the Minolta with 90 pixels or the Olympus with 120 pixels.

12. A: Cool, you're going to Barcelona in November?
    B: No, I'm going to Barcelona in January. I'm off to Vienna in November.

B.3 Block 3

1. A: OK, you want to catch the 19.30 train to London, then?
   B: No, I want to catch the 19.30 to Manchester and then the 20.30 to London.

2. A: You're going to see Amanda tomorrow, right?
   B: No, I'll see Amanda on Monday, I'm seeing Norma tomorrow.

3. A: Cool, you're going to Vienna in November?
   B: No, I'm off to Barcelona in November. I'm going to Vienna in January.

4. A: Are you interested in the Minolta camera with 90 mega pixels?
   B: No, I want either the Olympus with 120 pixels or the Minolta with 90 pixels.

5. A: So are you mailing me the magazines?
   B: No, I'm handing you the magazines. I'm mailing you the manuscripts.

6. A: Where are these mangoes from? Gautemala?
   B: No, the mangoes are from Australia. The limes are from Gautemala.

7. A: Cool, you're going to Barcelona in November?
   B: No, I'm off to Vienna in November. I'm going to Barcelona in January.

8. A: So, are you handing me the magazines?
   B: No, I'm handing you the manuscripts. I'm mailing you the magazines.

9. A: OK, you want to catch the 20.30 train to Manchester, then?
   B: No, I want to catch the 20.30 to London and then the 21.30 to Manchester.

10. A: Where are these mangoes from? Australia?
    B: No, the mangoes are from Gautemala. The limes are from Australia.

11. A: You're going to see Norma on Monday, right?
    B: No, I'm seeing Norma tomorrow, I'll see Amanda on Monday.

12. A: Are you interested in the Minolta camera with 120 mega pixels?
    B: No, I want either the Olympus with 120 pixels or the Minolta with 90 pixels.

B.4 Block 4

1. A: Where are these mangoes from? Australia?
   B: No, the limes are from Australia. The mangoes are from Gautemala.

2. A: So, are you mailing me the manuscripts?
   B: No, I'm mailing you the magazines. I'm handing you the manuscripts.

3. A: OK, you want to catch the 21.30 train to Manchester, then?
   B: No, I want to catch the 20.30 to Manchester and then the 21.30 to London.

4. A: Are you interested in the Olympus camera with 120 mega pixels?
   B: No, I want either the Minolta with 120 pixels or the Olympus with 90 pixels.

5. A: Cool, you're going to Vienna in November?
   B: No, I'm going to Vienna in January. I'm off to Barcelona in November.

6. A: You're going to see Norma on Monday, right?
   B: No, I'll see Amanda on Monday, I'm seeing Norma tomorrow.

7. A: Are you interested in the Minolta camera with 90 mega pixels?
   B: No, I want either the Minolta with 120 pixels or the Olympus with 90 pixels.

8. A: So, are you handing me the manuscripts?
   B: No, I'm handing you the magazines. I'm mailing you the manuscripts.

9. A: You're going to see Amanda tomorrow, right?
   B: No, I'm seeing Norma tomorrow, I'll see Amanda on Monday.

10. A: OK, you want to catch the 20.30 train to Manchester, then?
    B: No, I want to catch the 19.30 to Manchester and then the 20.30 to London.

11. A: Where are these mangoes from? Gautemala?
    B: No, the limes are from Gautemala. The mangoes are from Australia.

12. A: Cool, you're going to Vienna in January?
    B: No, I'm going to Vienna in November. I'm off to Barcelona in January.

Appendix C

Existing Corpus Annotations

This appendix gives details about the existing annotations of the Switchboard corpus.

C.1 Penn Treebank POS and Syntax

The Penn Treebank aimed to establish a common standard for annotating diverse corpora with Part of Speech (POS) information and syntactic structure in English, and included Switchboard. The full set of 33 part-of-speech tags used in the NXT version of the corpus is listed in Table C.1. A detailed description of the POS-tagging guidelines can be found in Santorini (1990), as well as additional notes on adapting the standards for Switchboard in LDC (n.d.).

Syntactic parsing was carried out using the Penn Treebank II standards. The full set of 24 phrase-level tags, and 21 function tags used in the NXT version of the corpus is listed in Tables C.2 and C.3. A detailed description of the tagset and annotation guidelines can be found in Bies, Ferguson & MacIntyre (1995) as well a description of additions to the original standards for the Switchboard corpus in Taylor (1996).

1.	BES	*'s* as form of *BE*	18.	PRP$	Possessive pronoun
2.	CC	Coordinating conjunction	19.	RB	Adverb
3.	CD	Cardinal number	20.	RBR	Adverb, comparative
4.	DT	Determiner	21.	RP	Particle
5.	EX	Existential *there*	22.	TO	infinitival *to*
6.	IN	Preposition/ subordinating conjunction	23.	UH	Interjection, filler, discourse marker
7.	JJ	Adjective	24.	VB	Verb, base form
8.	JJR	Adjective, comparative	25.	VBD	Verb, past tense
9.	JJS	Adjective, superlative	26.	VBG	Verb, gerund/ present participle
10.	MD	Modal	27.	VBN	Verb, past participle
11.	NN	Noun, singular or mass	28.	VBP	Verb, non-3rd ps. sing. present
12.	NNP	Proper noun, singular	29.	VBZ	Verb, 3rd ps. sing. present
13.	NNPS	Proper noun, plural	30.	WDT	*wh*-determiner
14.	NNS	Noun, plural	31.	WP	*wh*-pronoun
15.	PDT	Predeterminer	32.	WRB	*wh*-adverb
16.	POS	Possessive ending	33.	XX	Partial word, POS unclear
17.	PRP	Personal pronoun			

Table C.1: Treebank Part-Of-Speech tags used the NXT version of Switchboard

1.	ADVP	Adverb Phrase
2.	CONJP	Conjunction Phrase
3.	EDITED	Reparandum in disfluency
4.	FRAG	Fragment
5.	INTJ	Interjection, for words tagged UH
6.	IP	Interruption point in disfluency
7.	NAC	Not a constituent
8.	NP	Noun Phrase
9.	PP	Prepositional Phrase
10.	PRN	Parenthetical
11.	PRT	Particle, for words tagged RP
12.	QP	Quantifier Phrase
13.	RM	Reparandum in disfluency
14.	RS	Restart after disfluency
15.	S	Simple declarative clause
16.	SBAR	Clause introduced by a (possibly empty) subordinating conjunction
17.	SBARQ	Direct question introduced by a *wh*-word or *wh*-phrase
18.	SQ	Inverted *yes/no* question, or main clause of a *wh*-question
19.	TYPO	Speech Error
20.	UCP	Unlike Coordinated Phrase
21.	VP	Verb Phrase
22.	WHADVP	*Wh*-Adverb Phrase
23.	WHNP	*Wh*-Noun Phrase
24.	X	Unknown, uncertain or unbracketable

Table C.2: Treebank Phrase level tags used the NXT version of Switchboard

1.	ADV	Adverbial (other than ADVP or PP)
2.	DIR	Direction
3.	IMP	Imperative
4.	LOC	Locative
5.	LOC,PRD	Locative predicate
6.	MNR	Manner
7.	NOM	Nominal (on relatives and gerunds)
8.	NOM,TPC	Topicalised Nominal
9.	PRD	Predicate (other than VP)
10.	PRD,PRP	Purpose or reason predicate
11.	PRD,UNF	Unfinished Predicate
12.	PRP	Purpose or reason
13.	PRP,TPC	Topicalised purpose or reason
14.	PUT	Locative complement of *put*
15.	SBJ	Surface subject
16.	SBJ,UNF	Unfinished Surface Subject
17.	SEZ	Reported speech
18.	TMP	Temporal
19.	TMP,UNF	Unfinished Temporal
20.	TPC	Topicalised
21.	UNF	Unfinished

Table C.3: Treebank Function tags used the NXT version of Switchboard

C.2 Dialog Acts

Dialog acts were based on the DAMSL set of tags (Core & Allen 1997). However, some tags were grouped so that 42 different tags were used. They were defined so as to group utterances according to discourse purpose, discourse distribution and prosodic features. Dialog acts were annotated over *slash units*, conversational units marked by human labellers on the basis of pauses and discourse information (see Mateer & Taylor 1995). These units corresponded about 80% of the time to Treebank sentences, however, some slash units crossed sentence boundaries and some sentences contained more than one unit. The full list of 43 dialog act tags is given in Table C.4. A detailed description of the annotation guidelines can be found in Jurafsky, Shriberg & Biasca (1997).

C.3 Information Status

In addition to the broad three-way classification between *old*, *mediated* and *new* entities described in the main text, annotators could mark *old* and *mediated* subtypes (Nissim 2003, Nissim et al. 2004). These are listed in Tables C.5 and C.6. Annotators were given a decision tree in cases where more than one category applied. In the experiment data set, *infotype* was included as a separate feature. part, poss and func_value were grouped with set; and aggregation and situation with event, as these were felt to behave similarly, and some subtypes were very infrequent.

	NXT tag	DAMSL tag	Description	Freq.
1.	abandon	%-	Adandoned or Turn-Exit	7330
2.	acknowledge	bk	Response Acknowledgement	809
3.	affirm	na,ny^ e	Affirmative non-yes answers	479
4.	agree	aa	Agree/Accept	6356
5.	ans_dispref	arp,nd	Dispreferred answers	137
6.	answer	no	Other answers	193
7.	apology	fa	Apology	42
8.	apprec	ba	Appreciation	2662
9.	backchannel	b	Acknowledge (Backchannel)	19438
10.	backchannel_q	bh	Backchannel in question form	645
11.	close	fc	Conventional-closing	1444
12.	commit	oo,cc,co	Offers, Options Commits	69
13.	completion	^ 2	Collaborative Completion	353
14.	decl_q	qw^ d	Declarative Wh-Question	57
15.	directive	ad	Action-directive	420
16.	downplay	bd	Downplayer	39
17.	excluded	@	Slash unit excluded - bad segmentation	638
18.	hedge	h	Hedge	707
19.	hold	^ h	Hold before answer/agreement	340
20.	maybe	aap/am	Maybe/Accept-part	61
21.	neg	ng,nn^ e	Negative non-no answers	162
22.	no	nn	No answers	738
23.	open	fp	Conventional-opening	130
24.	open_q	qo	Open-Question	403
25.	opinion	sv	Statement-opinion	16553
26.	or	qrr	Or-Clause	111
27.	other	o,fo,bc,by,fw	Other	468
28.	quote	^ q	Quotation	579
29.	reject	ar	Reject	220
30.	repeat	b^ m	Repeat-phrase	382
31.	repeat_q	br	Signal-non-understanding	147
32.	rhet_q	qh	Rhetorial-Questions	357
33.	self_talk	t1	Self-Talk	43
34.	statement	sd	Statement-non-opinion	46151
35.	sum	bf	Summarize/Reformulate	585
36.	tag_q	^ g	Tag-Question	22
37.	thank	ft	Thanking	35
38.	third_pty	t3	3rd-party-talk	54
39.	uninterp	%	Uninterpretable	1578
40.	wh_q	qw	Wh-Question	1185
41.	yes	ny	Yes answers	1672
42.	yn_decl_q	qy^ d	Declarative Yes-No-Question	740
43.	yn_q	qy	Yes-No-Question	2816
			Total:	**117350**

Table C.4: Shriberg et al.'s (1998) Dialog act types, by NXT and original name

1.	ident	Anaphoric reference to a previously mentioned entity, e.g. I met *M.* **He**'s a nice guy"	12894
2.	relative	Relative pronoun	1600
3.	generic	Generic pronoun, e.g. "in holland **they** put mayo on chips"	3382
4.	ident_generic	Generic possesive pronoun, e.g. "in holland they put mayo on **their** chips"	2568
5.	general	"I" and "you"	10920
6.	event	Reference to a previously mentioned VP, e.g. "I like *going to the mountains.* Yeah, I like **it** too"	2544
7.	none	Sub-category not specified	1367
		Total:	**35299**

Table C.5: Nissim et al.'s (2004) Old subtypes

1.	bound	Bound pronoun, e.g. "*everyone* likes **his** job"	677
2.	general	Generally known, e.g. " the sun"	3354
3.	event	Relates to a previously mentioned VP, e.g. "We were *travelling around Yucatan*, and **the bus** was really full"	479
4.	aggregation	Reference to previously mentioned co-ordinated NPs, e.g. *John... Ann...* **they**"	943
5.	func_value	Refers to the value of a previously mentioned function, e.g. "in ... *centigrade* ... if it's between **zero** and **ten** it's cold"	79
6.	set	Subset, superset, or member of the same set as a previously mentioned entity	13645
7.	part	Part-whole relation for physical objects, both intra- and inter-phrasal, e.g. "when I come *home* ... my dog greets me at **the door**"	468
8.	poss	Intra-phrasal possessive relation (pre- and post-nominal) that is not *part*	1754
9.	situation	Part of a situation set up by a previous entity, e.g. "*capital punishment* ... **the exact specifications**"	1566
10.	none	Sub-category not specified	813
		Total:	**23816**

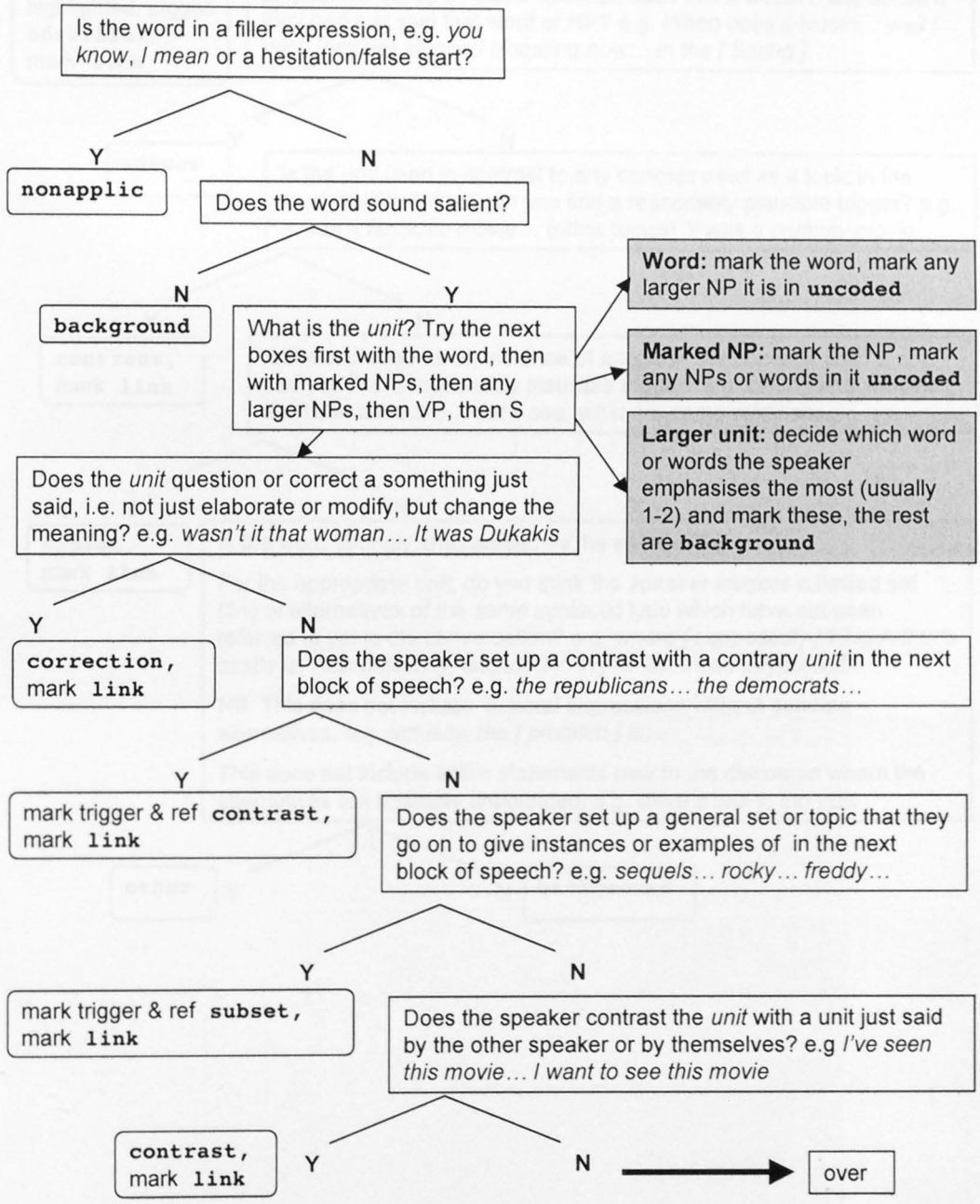
Table C.6: Nissim et al.'s (2004) Mediated subtypes

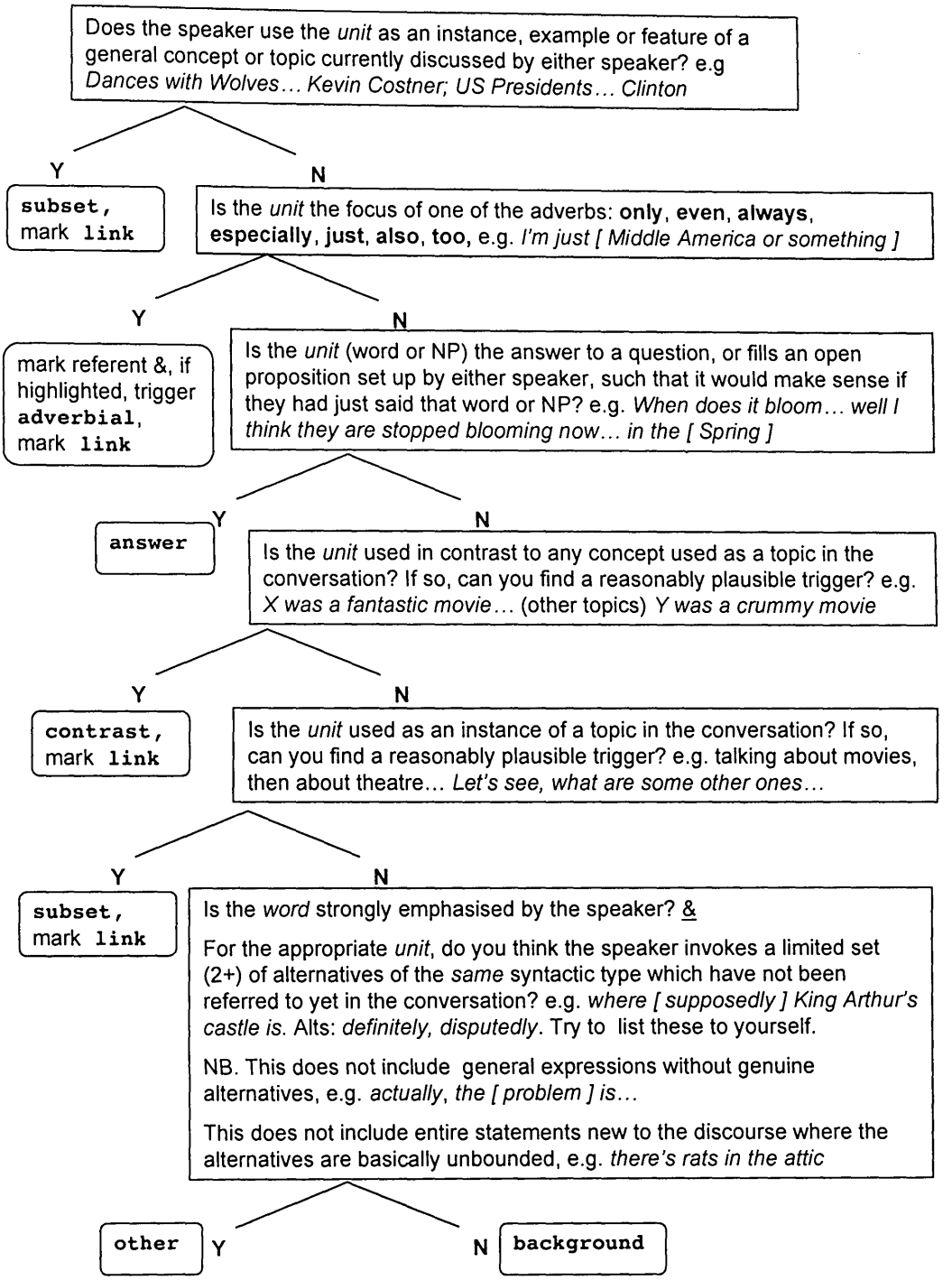
C.4 Phone and Syllable Alignment

Phone and syllable alignments were derived automatically using the Sonic speech recognition system (Pellom 2001) by Jason Brenier. Firstly phones were automatically aligned with the MS-State transcript using an existing lexicon of Switchboard. Another lexicon was used to group the phones into syllables and mark primary and secondly stress information. This technology is reasonably mature and error rates low enough that the data could be used without extensive manual checking. However, for short, disfluent words, stress, and in some cases, syllable information could not be determined. There were also a very small number of out-of-vocabulary items.

Appendix D

Kontrast Decision Tree




Does the speaker use the *unit* as an instance, example or feature of a general concept or topic currently discussed by either speaker? e.g *Dances with Wolves... Kevin Costner; US Presidents... Clinton*
Y
N
subset, mark link
Is the *unit* the focus of one of the adverbs: **only**, **even**, **always**, **especially**, **just**, **also**, **too**, e.g. *I'm just [Middle America or something]*
Y
N
mark referent &, if highlighted, trigger **adverbial**, mark **link**
Is the *unit* (word or NP) the answer to a question, or fills an open proposition set up by either speaker, such that it would make sense if they had just said that word or NP? e.g. *When does it bloom... well I think they are stopped blooming now... in the [Spring]*
Y
N
answer
Is the *unit* used in contrast to any concept used as a topic in the conversation? If so, can you find a reasonably plausible trigger? e.g. *X was a fantastic movie...* (other topics) *Y was a crummy movie*
Y
N
contrast, mark link
Is the *unit* used as an instance of a topic in the conversation? If so, can you find a reasonably plausible trigger? e.g. talking about movies, then about theatre... *Let's see, what are some other ones...*
Y
N
subset, mark link
Is the *word* strongly emphasised by the speaker? &
For the appropriate *unit*, do you think the speaker invokes a limited set (2+) of alternatives of the *same* syntactic type which have not been referred to yet in the conversation? e.g. *where [supposedly] King Arthur's castle is.* Alts: *definitely, disputedly*. Try to list these to yourself.
NB. This does not include general expressions without genuine alternatives, e.g. *actually, the [problem] is...*
This does not include entire statements new to the discourse where the alternatives are basically unbounded, e.g. *there's rats in the attic*
other
Y
N
background

Appendix E

Features Used in Chapter 6

E.1 Glossary of Features

A list of the features used in the studies described in Chapter 6. Extraction of these features is described in Chapter 5. In addition to the feature name, a brief description of the feature is given, along with the possible values of that feature. For continuous variables, the type of number value it can take is given, i.e. *whole* (whole number) or *float* (other rational number).

E.1.1 Discourse Semantic

Feature	Description	Values
kon_stat	whether the word is a kontrast, i.e. not background, at the word level or np level	konword, konnp, backgd
is_kon	whether the word is the head of a kontrast, i.e. either a konword, or both a konnp and the head of its constituent	kon, backgd
kon_type	the kontrast type of the word	correct, contrast, subset_kon, adverbial_kon, answer, other_kon, backgd
kon_bound	whether the word is not in a kon, in the middle of a kon phrase, or ends a kon	notinkon, inkon, konbound

Feature	Description	Values
dist_trig	how many words to the trigger of the kon (only for correct, contrast, subset_kon and adverbial_kon types)	*whole*
distKon_cl	how many words since the last kon, or the beginning of the clause	*whole*
numKon_ph	how many distinct kons, i.e. not just words in the same konnp, so far in the phrase	*whole*
next_kon	whether the next word is kon or backgd	nextkon, nextbackgd
info_stat	the information status of the word (only NPs classified)	old, med, new, noinf
info_type	the information status subtype of the word	ident_inf, rel_inf, generic_inf, general_inf, event_inf, bound_inf, set_inf, no_inftype
info_bound	whether the word is not classified for infostat, in a marked NP, or the last word in a marked NP (lowest level of recursive structure)	notininf, ininf, infbound
dist_coref	how many words to the last mention of the entity (only words classified old_ident and old_rel)	*whole*
next_info	the information status of the next word	nextold, nextmed, nextnew
dial_act	the dialog act of the word (grouped into the 3 most common types and other)	statement, opinion, question, other
disfl	whether the word is classfied as disfluent, in a repair, or not disfluent	reparandum, repair, notdisfl
tr_stat	the theme/rheme status of the word (only one conversation)	theme, rheme
tr_place	whether the theme/rheme appears before or after its pair	first, second
tr_order	whether the information unit occurs in theme/rheme or rheme/theme order	t-r, r-t
tr_N	whether the word carries the last nuclear accent in the theme or rheme phrase	nuc, not

E.1.2 Syntactic

Feature	Description	Values
clause_type	the type of clause the word is in	main_cl, comp_cl, rel_cl, adv_cl, paren_cl
posWd_cl	the position of the word in the clause	*whole*
numWd_cl	the number of words in the clause in total	*whole*
propWd_cl	the position of the current word relative to the number of words in the clause	*float*
cl_bound	whether or not the word is the last word is the clause	incl, clbound
cl_haskon	whether the clause has at least one kon in it	konincl, nokonincl
constit_type	the type of constituent the word is in	subj, pred, obj, adjunct
posWd_cns	the position of the word in the constituent	*whole*
numWd_cns	the number of words in the constituent in total	*whole*
propWd_cns	the position of the current word relative to the number of words in the constituent	*float*
cns_bound	whether or not the word is the last word in the constituent	incns, cnsbound
cns_hasKon	whether the constituent has at least one kon in it	konincns, nokonincns
is_cnsHead	whether the word is the head of its constituent	nothead_cns, head_cns
POS_gp	the part-of-speech of the word, by major group	NN, VB, PR, DT, JJ, RB, XX

E.1.3 Phrasal

Feature	Description	Values
break_type	whether the word is followed by a 3 break, 4 break or no break	nobrk, 3brk, 4brk
is_break	whether the word is followed by a break, either 3brk or 4brk	nobrk, brk
inph_type	the type of phrase the word is in	minor, major
posWd_ph	the position of the word in the phrase	*whole*
numWd_ph	the number of words in the phrase in total	*whole*
propWd_ph	the position of the current word relative to the number of words in the phrase	*float*

wd_lastPh	the number of words in the preceding phrase	*whole*
wdRel_lastPh	the position of the word in the current phrase relative to the number of words in the preceding phrase	*float*
posSyl_ph	the position of the syllable in the phrase	*whole*
numSyl_ph	the number of syllables in the phrase in total	*whole*
propSyl_ph	the position of the current syllable relative to the number of syllables in the phrase	*float*
syls_lastPh	the number of syllables in the preceding phrase	*whole*
sylRel_lastPh	the position of the syllable in the current phrase relative to the number of syllables in the preceding phrase	*float*
posPho_ph	the position of the phone in the phrase	*whole*
numPho_ph	the number of phones in the phrase in total	*whole*
propPho_ph	the position of the current phone relative to the number of phones in the phrase	*float*
npmean_ph	mean pitch in the phrase, normalised as a percentage of the speaker's overall range in logged Hz	*float*
nimean_ph	mean intensity in the phrase relative to the mean intensity of all phrases for that speaker	*float*
t_ph	duration of the phrase up to and including the word	*float*
t_lastPh	duration of the previous phrase	*float*
tRel_lastPh	duration of the phrase so far relative to the duration of the previous phrase	*float*
spRate_wd	the total number of words in the phrase relative to the phrase duration	*float*
spRate_syl	the total number of syllables in the phrase relative to the phrase duration	*float*

E.1.4 Accentual

Feature	Description	Values
acc_type	the full accent type	Q, A, PN, N, noacc
accq_gp	the accent group, counting Q, A as accents and PN, N as nuclear	accq, nuc, noaccq
accnq_gp	the accent group, counting A as accents and PN, N as nuclear	accnq, nuc, noaccnq
acc_stat	accent status by position, i.e. pre-nuclear, nuclear or post-nuclear (includes Q as accented)	pre, nuc, post, noacc
is_accq	whether the word is an accent, i.e. Q, A, PN, N	anyaccq, noaccq
is_accnq	whether the word is an accent, i.e. A, PN, N	anyaccnq, noaccnq
is_nuc	whether the word is nuclear, i.e. PN, N	nuc, notnuc
naccH	pitch at the marked accent peak, normalised as a percentage of the speaker's overall range in logged Hz	*float*
naccH_time	time of the marked accent peak, normalised relative to the stressed syllable of the word	*float*
naccL_time	time of the pitch minimum in the word, normalised relative to the stressed syllable of the word (only if occurs before naccH_time)	*float*
accsPh_inc	number of accents in the phrase so far, including the current word	*whole*
accsPh_exc	number of accents in the phrase so far, excluding the current word	*whole*
num_accPh	number of accents in the phrase in total	*whole*
accq_dist	number of words since the last accent (including Q)	*whole*
accnq_dist	number of words since the last accent (excluding Q)	*whole*
numSyl_wd	number of syllables in the word	*whole*
pos_strSyl	position of the stressed syllable in the word	*whole*
numPho_wd	number of phones in the word	*whole*

E.1.5 Word level Acoustic

Feature	Description	Values
npmin_wd	minimum pitch in the word, normalised as a percentage of the speaker's overall range in logged Hz	*float*
npmax_wd	maximum pitch in the word, normalised as a percentage of the speaker's overall range in logged Hz	*float*
npmean_wd	mean pitch in the word, normalised as a percentage of the speaker's overall range in logged Hz	*float*
npquan_wd	pitch at the 0.5 quantile, normalised as a percentage of the speaker's overall range in logged Hz	*float*
nprange_wd	the pitch range in the word, i.e. npmax_wd - npmin_wd	*float*
npqrange_wd	the inter-quantile pitch range in the word, i.e. 0.75 - 0.25, normalised as a percentage of the speaker's overall range in logged Hz	*float*
nimean_wd	the mean intensity in the word, relative to the mean intensity of all words for that speaker	*float*
dur_relPho	the duration of the word relative to the number of phones in the word	*float*
dur_relSyl	the duration of the word relative to the number of syllables in the word	*float*
prom_wd	an approximate measure of prominence for the word, calculated by: $(2 * dur\_relSyl + nqrange\_wd + npquan\_wd + (nimean\_wd - 5))/10$	*float*

E.2 Features Used in Each Model

Features tested in the models described in Chapter 6. Significant features are ticked.

E.2.1 Phrase Break Prediction [DV: is_break]

Models in P1: P (phrasal); S (semantic/syntactic); A (accentual); W (acoustic word).

Phrasal Features

	CART			Regr.				CART			Regr.		
Feature	**P**	**SP**	**-A**	**P**	**SP**	**-A**	**Feature**	**P**	**SP**	**-A**	**P**	**SP**	**-A**
posWd_ph				√	√	√	npmean_ph				√		
posSyl_ph	√	√					nimean_ph						
posPho_ph				√	√	√	spRate_syl	√	√	√	√	√	√
t_ph	√	√	√	√	√	√	disfl				√	√	
wd_lastPh				√	√	√	wdRel_lastPh						
syls_lastPh							sylRel_lastPh				√	√	√
t_lastPh	√		√				tRel_lastPh						

Accentual Features

	CART		Regr.			CART		Regr.	
Feature	**P**	**SP**	**P**	**SP**	**Feature**	**P**	**SP**	**P**	**SP**
is_accq	√	√	√	√	accq_dist				
accsPh_inc	√	√	√	√					

Semantic/Syntactic Features

	CART			Regr.				CART			Regr.		
Feature	**S**	**SP**	**-A**	**S**	**SP**	**-A**	**Feature**	**S**	**SP**	**-A**	**S**	**SP**	**-A**
kon_stat							clause_type				√	√	
kon_type							numWd_cl	√	√	√	√	√	√
kon_bound	√			√	√	√	cl_bound				√	√	√
numKon_ph		√		√			propWd_cl	√	√	√			
next_kon	√	√	√				constit_type			√	√	√	√
info_stat							numWd_cns			√			
info_type				√	√	√	cns_bound				√	√	√
info_bound				√	√	√	propWd_cns		√	√	√	√	√
next_info	√	√	√	√	√	√	cns_hasKon						
dial_act							is_cnsHead	√			√	√	√
disfl				√	√		POS_gp	√		√	√	√	√

Word Acoustic Features (only in regression model)

Feature	**SPW**	**Feature**	**SPW**	**Feature**	**SPW**	**Feature**	**SPW**
npmin_wd		npmean_wd		nprange_wd		nimean_wd	
npmax_wd		npquan_wd	√	npqrange_wd	√	dur_relSyl	√

E.2.2 Plain v Nuclear Accent Prediction [DV: is_nuc (noaccq excl)]

Models in A1: P (phrasal); S (semantic/syntactic); W (acoustic word).

PSW model excludes phrasal position features.

Phrasal Position Features

Feature	CT	Reg	Feature	CT	Reg	Feature	CT	Reg
numWd_ph		√	posWd_ph			propWd_ph		
numSyl_ph	√		posSyl_ph		√	propSyl_ph	√	
numPho_ph	√	√	posPho_ph			propPho_ph	√	√
t_ph	√	√	is_break		√	accsPh_exc		√

Other Phrasal Features

	CART		Regr.			CART		Regr.	
Feature	P	SPW	P	SPW	Feature	P	SPW	P	SPW
npmean_ph					nimean_ph	√			
disfl					numSyl_wd		√	√	√
spRate_syl	√	√			accq_dist				

Semantic/Syntactic Features

	CART		Regr.			CART		Regr.	
Feature	S	SPW	S	SPW	Feature	S	SPW	S	SPW
kon_stat	√	√	√	√	clause_type				
kon_type					numWd_cl	√	√	√	√
numKon_ph					propWd_cl	√	√	√	√
distKon_cl	√	√			constit_type	√		√	√
disfl					numWd_cns	√			
info_stat	√				propWd_cns	√		√	√
info_type	√				cns_hasKon				
POS_gp	√	√			is_cnsHead				

Word Acoustic Features (only in PSW model)

Feature	CT	Reg	Feature	CT	Reg	Feature	CT	Reg
npmin_wd			npquan_wd		√	nimean_wd		
npmax_wd			nprange_wd			dur_relSyl	√	√
npmean_wd	√		npqrange_wd	√	√			

E.2.3 Plain v No Accent Prediction [DV: is_accq (nuc excl)]

Models in A2: P (phrasal); S (semantic/syntactic); W (acoustic word).

+W model is PS+W. PSW has the same S features as the PS model.

Phrasal Features

	CART			Regr.				CART			Regr.		
Feature	**P**	**PS**	**+W**	**P**	**PS**	**+W**	**Feature**	**P**	**PS**	**+W**	**P**	**PS**	**+W**
numWd_ph					√	√	t_ph	√	√	√	√	√	√
posWd_ph							is_break					√	√
propWd_ph	√	√	√	√	√	√	accsPh_exc	√	√	√	√	√	√
numSyl_ph							npmean_ph				√	√	√
posSyl_ph				√	√		nimean_ph	√	√	√			
propSyl_ph							disfl						
numPho_ph	√						numSyl_wd	√	√	√	√	√	√
posPho_ph							spRate_syl	√	√	√	√	√	
propPho_ph	√		√				accq_dist				√	√	√

Semantic/Syntactic Features

	CART		Regr.			CART		Regr.	
Feature	**S**	**PS**	**S**	**PS**	**Feature**	**S**	**PS**	**S**	**PS**
kon_stat	√		√	√	clause_type	√	√		
kon_type					numWd_cl	√	√		
numKon_ph	√	√	√	√	propWd_cl	√	√		
distKon_cl		√			constit_type			√	√
disfl					numWd_cns				
info_stat	√	√			propWd_cns	√	√	√	√
info_type	√				cns_hasKon	√		√	
POS_gp	√	√	√	√	is_cnsHead	√			

Word Acoustic Features (only in PSW model)

Feature	**CT**	**Reg**	**Feature**	**CT**	**Reg**	**Feature**	**CT**	**Reg**
npmin_wd			npquan_wd	√	√	nimean_wd	√	√
npmax_wd			nprange_wd			dur_relSyl	√	√
npmean_wd			npqrange_wd					

E.2.4 Nuclear v No Accent Prediction [DV: is_nuc (accq excl)]

Models in A2: P (phrasal); S (semantic/syntactic); W (acoustic word).

PSW model has the same P & S features as the PS model.

Phrasal Features

	CART		Regr.			CART		Regr.	
Feature	**P**	**PS**	**P**	**PS**	**Feature**	**P**	**PS**	**P**	**PS**
numWd_ph			√	√	t_ph	√		√	√
posWd_ph			√	√	is_break			√	√
propWd_ph					accsPh_exc	√	√	√	√
numSyl_ph	√	√			npmean_ph			√	√
posSyl_ph			√	√	nimean_ph				
propSyl_ph	√				disfl				
numPho_ph					numSyl_wd	√			
posPho_ph					spRate_syl			√	√
propPho_ph	√	√	√	√	accq_dist			√	√

Semantic/Syntactic Features

	CART		Regr.			CART		Regr.	
Feature	**S**	**PS**	**S**	**PS**	**Feature**	**S**	**PS**	**S**	**PS**
kon_stat	√	√	√	√	clause_type				
kon_type					numWd_cl				
numKon_ph			√	√	propWd_cl	√		√	
distKon_cl					constit_type				
disfl					numWd_cns				
info_stat					propWd_cns			√	
info_type	√				cns_hasKon	√	√		
POS_gp	√	√	√		is_cnsHead				

Word Acoustic Features (only in PSW model)

Feature	**CT**	**Reg**	**Feature**	**CT**	**Reg**	**Feature**	**CT**	**Reg**
npmin_wd		√	npquan_wd	√	√	nimean_wd		√
npmax_wd			nprange_wd	√		dur_relSyl	√	√
npmean_wd	√		npqrange_wd					

E.2.5 Accent Prediction [DV: is_accq]

Models in A3: P (phrasal); S (semantic/syntactic); W (acoustic word).

+W model is PS+W.

Phrasal Features

	CART			Regr.				CART			Regr.		
Feature	**P**	**PS**	**+W**	**P**	**PS**	**+W**	**Feature**	**P**	**PS**	**+W**	**P**	**PS**	**+W**
numWd_ph				√	√	√	t_ph	√	√	√	√	√	√
posWd_ph							is_break				√	√	√
propWd_ph	√	√	√	√	√	√	accsPh_exc	√	√	√	√	√	√
numSyl_ph	√	√	√				npmean_ph	√			√	√	√
posSyl_ph				√	√		nimean_ph	√	√	√			
propSyl_ph							disfl						
numPho_ph	√						numSyl_wd	√	√	√	√	√	√
posPho_ph							spRate_syl	√	√	√	√	√	√
propPho_ph	√			√	√	√	accq_dist				√	√	√

Semantic/Syntactic Features

	CART			Regr.				CART			Regr.		
Feature	**S**	**PS**	**+W**	**S**	**PS**	**+W**	**Feature**	**S**	**PS**	**+W**	**S**	**PS**	**+W**
kon_stat	√	√	√	√	√	√	clause_type	√	√	√			
kon_type				√			numWd_cl						
numKon_ph				√	√		propWd_cl	√					
distKon_cl							constit_type				√	√	
disfl							numWd_cns	√					
info_stat							propWd_cns				√	√	
info_type	√						cns_hasKon						
POS_gp	√	√	√	√	√	√	is_cnsHead						

Word Acoustic Features (only in PSW model)

Feature	**CT**	**Reg**	**Feature**	**CT**	**Reg**	**Feature**	**CT**	**Reg**
npmin_wd			npquan_wd		√	nimean_wd	√	√
npmax_wd			nprange_wd			dur_relSyl	√	√
npmean_wd			npqrange_wd	√				

E.2.6 Accent Group Prediction [DV: accq_gp]

Models in A3: P (phrasal); S (semantic/syntactic); W (acoustic word).

PSW model has same P & S features as PS model.

–P is the PS model without phrase proportion features.

Phrase Proportion Features

	CART		Regr.			CART		Regr.	
Feature	**P**	**PSW**	**P**	**PSW**	**Feature**	**P**	**PSW**	**P**	**PSW**
numWd_ph			√	√	propWd_ph	√	√	√	
numSyl_ph					propSyl_ph				
numPho_ph			√	√	propPho_ph	√	√	√	√
is_break			√	√					

Other Phrasal Features

	CART			Regr.				CART			Regr.		
Feature	**P**	**PS**	**-P**	**P**	**PS**	**-P**	**Feature**	**P**	**PS**	**-P**	**P**	**PS**	**-P**
posWd_ph			√	√			npmean_ph				√	√	√
posSyl_ph	√	√	√	√	√	√	nimean_ph	√					
posPho_ph							disfl						
t_ph	√	√	√	√	√	√	spRate_syl	√			√	√	√
accsPh_exc	√	√	√	√	√	√	numSyl_wd	√		√	√	√	√
accq_dist				√	√	√							

Semantic/Syntactic Features

	CART			Regr.				CART			Regr.		
Feature	**S**	**PS**	**-P**	**S**	**PS**	**-P**	**Feature**	**S**	**PS**	**-P**	**S**	**PS**	**-P**
kon_stat	√	√	√	√	√	√	clause_type	√	√	√			
kon_type							numWd_cl	√	√	√	√		√
numKon_ph	√			√	√		propWd_cl	√	√	√	√		√
distKon_cl	√						constit_type	√	√	√			√
disfl							numWd_cns	√		√			
info_stat							propWd_cns	√	√	√	√		√
info_type	√		√				cns_hasKon						
POS_gp	√	√	√	√	√	√	is_cnsHead						

Word Acoustic Features (only in PSW model)

Feature	**CT**	**Reg**	**Feature**	**CT**	**Reg**	**Feature**	**CT**	**Reg**
npmin_wd			npquan_wd	√	√	nimean_wd	√	√
npmax_wd			nprange_wd	√		dur_relSyl	√	√
npmean_wd	√		npqrange_wd		√			

Appendix F

Full Result Tables from Chapter 6

F.1 Parameter Estimates for P1

Table F.1: All parameter estimates for the full phrase prediction model in P1

Feat	Exp(B)	Sig	Wald (df)
kon_bound	-	.027	7.2 (2)
inkon	0.68	.018	5.6 (1)
konbound	1.05	.588	0.3 (1)
disfl	-	.001	14.5 (2)
repair	3.77	.001	11.1 (1)
notdisfl	0.74	.085	3.0 (1)
numWd_cl	1.03	.000	40.2 (1)
clbound	4.76	.000	241.1 (1)
head_cns	0.74	.000	12.3 (1)
cnsbound	4.04	.000	43.2 (1)
info_type	-	.000	50.9 (7)
ident_inf	1.98	.000	21.1 (1)
rel_inf	0.10	.000	17.4 (1)
generic_inf	1.05	.834	0.0 (1)
bound_inf	2.94	.008	6.9 (1)
set_inf	0.97	.812	0.1 (1)
general_inf	1.12	.477	0.5 (1)
event_inf	1.65	.003	8.9 (1)
next_info	-	.000	122.6 (3)
nextold	0.69	.000	22.7 (1)
nextmed	0.83	.015	5.9 (1)
nextnew	0.92	.413	0.7 (1)
propWd_cns	0.20	.000	26.7 (1)

Feat	Exp(B)	Sig	Wald (df)
POS_gp	-	.000	24.2 (6)
RB	1.14	.449	0.6 (1)
JJ	1.28	.171	1.9 (1)
PR	1.32	.148	2.1 (1)
VB	0.73	.069	3.3 (1)
NN	1.45	.019	5.5 (1)
DT	1.17	.408	0.7 (1)
cns_bound * constit_type	-	.000	45.3 (3)
cnsbound by adjunct	2.36	.000	31.5 (1)
cnsbound by obj	1.56	.002	9.7 (1)
cnsbound by subj	0.32	.000	25.3 (1)
t_ph	1.62	.000	299.6 (1)
sylRel_lastPh	0.93	.006	7.6 (1)
accsPh_inc	1.91	.000	76.4 (1)
anyaccq	1.70	.000	31.3 (1)
posPho_ph	0.85	.000	64.4 (1)
posWd_ph	0.31	.000	223.7 (1)
posWd_ph by spRate_syl	1.15	.000	202.5 (1)
posWd_ph by wd_lastPh	0.99	.001	10.2 (1)
constit_type * posWd_ph * propWd_cns	-	.002	14.7 (3)
adjunct by posWd_ph by propWd_cns	1.05	.014	6.0 (1)
obj by posWd_ph by propWd_cns	1.03	.090	2.9 (1)
subj by posWd_ph by propWd_cns	0.89	.000	14.3 (1)
npquan_wd	0.91	.000	44.3 (1)
dur_relSyl	1.97	.000	285.9 (1)
npqrange_wd	1.10	.000	16.5 (1)
Constant	0.06	.000	61.0 (1)

F.2 Parameter Estimates for A1

Table F.2: All parameter estimates for the plain v nuclear accent prediction model

Feat	Exp(B)	Sig	Wald (df)
kon_stat	-	.000	106.7 (2)
konnp	1.50	.000	17.8 (1)
konword	2.23	.000	106.5 (1)
numWd_cl	1.05	.000	23.1 (1)
propWd_cl	3.00	.000	29.2 (1)
propWd_cns	2.75	.000	67.3 (1)
numWd_cl by propWd_cl	0.94	.000	16.0 (1)
constit_type * propWd_cns	-	.000	60.5 (3)
adjunct by propWd_cns	1.55	.000	32.8 (1)
obj by propWd_cns	1.32	.001	12.0 (1)
subj by propWd_cns	0.69	.000	15.8 (1)
numSyl_wd	1.54	.000	95.0 (1)
spRate_syl	1.17	.000	30.7 (1)
npquan_wd	0.81	.000	42.0 (1)
dur_relSyl	2.18	.000	308.6 (1)
npqrange_wd	1.18	.000	46.6 (1)
npmean_ph by npquan_wd	1.02	.000	47.3 (1)
Constant	0.01	.000	203.4 (1)

F.3 Parameter Estimates for A2

Table F.3: All parameter estimates for the plain accent v no accent prediction model

Feat	Exp(B)	Sig	Wald (df)
kon_stat	-	.000	66.0 (2)
konnp	1.32	.211	1.6 (1)
konword	4.33	.000	45.5 (1)
constit_type	-	.000	21.7 (3)
adjunct	1.58	.013	6.1 (1)
obj	1.17	.353	0.9 (1)
subj	1.19	.481	0.5 (1)
POS_gp	-	.000	47.3 (6)
RB	1.11	.589	0.3 (1)
JJ	1.15	.493	0.5 (1)
PR	0.58	.007	7.4 (1)
VB	1.01	.949	0.0 (1)
NN	1.48	.039	4.3 (1)
DT	1.52	.038	4.3 (1)
kon_stat * numKon_ph	-	.001	14.6 (2)
konnp by numKon_ph	0.75	.163	1.9 (1)
konword by numKon_ph	0.51	.001	11.8 (1)
constit_type * propWd_cns	-	.000	30.5 (3)
adjunct by propWd_cns	0.56	.013	6.2 (1)
obj by propWd_cns	0.57	.012	6.2 (1)
subj by propWd_cns	1.20	.509	0.4 (1)
numWd_ph	0.91	.001	10.2 (1)
numSyl_wd	2.90	.000	222.3 (1)
t_ph	1.23	.000	29.2 (1)
npmean_ph	0.77	.000	117.5 (1)
propWd_ph	0.18	.000	21.2 (1)
brk	1.61	.006	7.7 (1)
accq_dist	1.77	.000	119.4 (1)
accsPh_exc	0.07	.000	139.6 (1)
accsPh_exc by numWd_ph	1.19	.000	34.5 (1)
numWd_ph by t_ph	0.99	.000	12.5 (1)
npquan_wd	1.43	.000	257.8 (1)
nimean_wd	1.63	.000	178.7 (1)
dur_relSyl	2.44	.000	258.5 (1)
Constant	0.00	.000	214.7 (1)

Table F.4: All parameter estimates for the nuclear accent v no accent prediction model

Feat	Exp(B)	Sig	Wald (df)
kon_stat	-	.000	189.0 (2)
konnp	3.82	.000	26.8 (1)
konword	21.45	.000	170.9 (1)
numKon_ph	0.59	.000	29.2 (1)
kon_stat * numKon_ph	-	.004	10.9 (2)
konnp by numKon_ph	1.03	.906	0.0 (1)
konword by numKon_ph	0.56	.001	10.2 (1)
posWd_ph	0.37	.000	71.0 (1)
numWd_ph	0.55	.000	197.6 (1)
accq_dist	1.88	.000	123.3 (1)
accsPh_exc	0.07	.000	158.0 (1)
brk	0.37	.000	57.2 (1)
numWd_ph by posSyl_ph	0.95	.000	75.0 (1)
accsPh_exc by numWd_ph	1.38	.000	130.2 (1)
numWd_ph by propPho_ph	3.29	.000	101.3 (1)
spRate_syl by t_ph	1.08	.000	262.7 (1)
accsPh_exc by npmean_ph	0.88	.000	51.9 (1)
npmin_wd	0.80	.000	62.8 (1)
npquan_wd	1.57	.000	221.3 (1)
nimean_wd	1.56	.000	94.5 (1)
dur_relSyl	2.22	.000	301.9 (1)
Constant	0.00	.000	152.6 (1)

F.4 Parameter Estimates for A3

Table F.5: All parameter estimates for the full accent type prediction model in A3

Accent			
Feat	**Exp(B)**	**Sig**	**Wald (df)**
Intercept	-	.000	211.8 (1)
konnp	1.52	.061	3.5 (1)
konword	2.92	.000	28.5 (1)
backgd	-	-	- (0)
XX	0.53	.491	0.5 (1)
RB	0.75	.029	4.8 (1)
JJ	0.85	.274	1.2 (1)
PR	0.56	.000	21.2 (1)
VB	0.73	.007	7.4 (1)
NN	0.95	.706	0.1 (1)
DT	-	-	- (0)
konnp * numKon_ph	0.82	.255	1.3 (1)
konword * numKon_ph	0.72	.032	4.6 (1)
backgd * numKon_ph	1.18	.044	4.1 (1)
brk	0.85	.203	1.6 (1)
nobrk	-	-	- (0)
numSyl_wd	2.28	.000	213.1 (1)
posSyl_ph	0.91	.138	2.2 (1)
accq_dist	1.34	.000	63.7 (1)
accsPh_exc	0.33	.000	101.9 (1)
npmean_ph	0.84	.000	85.7 (1)
propPho_ph	0.15	.000	32.1 (1)
numPho_ph	0.96	.000	27.6 (1)
propPho_ph * numPho_ph	1.08	.001	11.7 (1)
accsPh_exc * numWd_ph	1.01	.332	0.9 (1)
spRate_syl * t_ph	1.02	.003	9.1 (1)
npquan_wd	1.27	.000	188.1 (1)
nimean_wd	1.41	.000	128.3 (1)
dur_relSyl	2.01	.000	320.4 (1)
npqrange_wd	1.00	.797	0.1 (1)

Nuclear			
Feat	**Exp(B)**	**Sig**	**Wald (df)**
Intercept	-	.000	260.8 (1)
konnp	1.68	.014	6.0 (1)
konword	5.24	.000	81.1 (1)
backgd	-	-	- (0)
XX	1.51	.653	0.2 (1)
RB	0.86	.289	1.1 (1)
JJ	0.94	.724	0.1 (1)
PR	0.81	.146	2.1 (1)
VB	0.88	.327	1.0 (1)
NN	1.17	.253	1.3 (1)
DT	-	-	- (0)
konnp * numKon_ph	0.91	.529	0.4 (1)
konword * numKon_ph	0.69	.004	8.5 (1)
backgd * numKon_ph	0.87	.106	2.6 (1)
brk	1.82	.000	23.8 (1)
nobrk	-	-	- (0)
numSyl_wd	2.38	.000	222.3 (1)
posSyl_ph	0.61	.000	47.2 (1)
accq_dist	1.34	.000	62.3 (1)
accsPh_exc	0.24	.000	196.5 (1)
npmean_ph	0.77	.000	146.4 (1)
propPho_ph	1.81	.110	2.6 (1)
numPho_ph	0.89	.000	94.5 (1)
propPho_ph * numPho_ph	1.12	.000	20.9 (1)
accsPh_exc * numWd_ph	1.07	.000	33.7 (1)
spRate_syl * t_ph	1.06	.000	84.5 (1)
npquan_wd	1.36	.000	258.2 (1)
nimean_wd	1.50	.000	138.5 (1)
dur_relSyl	2.02	.000	333.1 (1)
npqrange_wd	1.04	.011	6.4 (1)

F.5 Parameter Estimates for K1

Table F.6: All parameter estimates for the kontrast prediction model in K1

Feat	Exp(B)	Sig	Wald (df)
info_stat	-	.000	35.3 (3)
old	0.78	.113	2.5 (1)
med	2.26	.000	32.4 (1)
new	0.94	.734	0.1 (1)
head_cns	0.29	.000	64.5 (1)
constit_type * propWd_cns	-	.000	55.1 (3)
adjunct by propWd_cns	0.62	.001	10.7 (1)
obj by propWd_cns	0.62	.000	12.6 (1)
subj by propWd_cns	0.92	.633	0.2 (1)
POS_gp	-	.000	196.1 (6)
RB	2.03	.097	2.8 (1)
JJ	8.31	.000	24.9 (1)
PR	2.29	.056	3.7 (1)
VB	2.66	.022	5.3 (1)
NN	6.70	.000	20.6 (1)
DT	2.47	.035	4.5 (1)
propWd_cl	3.10	.000	117.3 (1)
info_stat * propSyl_ph	-	.000	24.1 (3)
old by propSyl_ph	0.81	.273	1.2 (1)
med by propSyl_ph	0.47	.000	18.1 (1)
new by propSyl_ph	2.31	.000	13.8 (1)
head_cns by propSyl_ph	2.72	.000	25.5 (1)
accq_gp * POS_gp	-	.000	38.6 (12)
accq by RB	1.14	.881	0.0 (1)
accq by JJ	10.44	.019	5.5 (1)
accq by PR	9.67	.012	6.3 (1)
accq by VB	2.11	.231	1.4 (1)
accq by NN	2.17	.149	2.1 (1)
accq by DT	6.98	.069	3.3 (1)
nuc by RB	1.20	.824	0.1 (1)
nuc by JJ	4.32	.121	2.4 (1)
nuc by PR	21.36	.001	11.1 (1)
nuc by VB	6.53	.001	10.4 (1)
nuc by NN	1.03	.946	0.0 (1)
nuc by DT	3.07	.284	1.1 (1)

Feat	Exp(B)	Sig	Wald (df)
constit_type * prom_wd	-	.000	35.2 (3)
adjunct by prom_wd	1.19	.015	5.9 (1)
obj by prom_wd	1.16	.024	5.1 (1)
subj by prom_wd	1.13	.190	1.7 (1)
accq_gp * POS_gp * prom_wd	-	.002	30.3 (12)
accq by RB by prom_wd	1.92	.248	1.3 (1)
accq by JJ by prom_wd	0.38	.120	2.4 (1)
accq by PR by prom_wd	0.50	.206	1.6 (1)
accq by VB by prom_wd	1.75	.157	2.0 (1)
accq by NN by prom_wd	0.77	.429	0.6 (1)
accq by DT by prom_wd	0.49	.280	1.2 (1)
nuc by RB by prom_wd	3.35	.019	5.5 (1)
nuc by JJ by prom_wd	1.14	.823	0.1 (1)
nuc by PR by prom_wd	0.43	.102	2.7 (1)
nuc by VB by prom_wd	1.64	.165	1.9 (1)
nuc by NN by prom_wd	1.89	.015	5.9 (1)
nuc by DT by prom_wd	1.00	.994	0.0 (1)
Constant	0.12	.000	26.0 (1)

Table F.7: All parameter estimates for the kontrast prediction model on nuclear accents

Feat	Exp(B)	Sig	Wald (df)
info_stat	-	.000	31.3 (3)
old	0.61	.000	13.8 (1)
med	1.19	.070	3.3 (1)
new	1.87	.000	23.7 (1)
head_cns	0.32	.003	8.8 (1)
constit_type * propWd_cns	-	.000	26.7 (3)
adjunct by propWd_cns	0.47	.002	10.0 (1)
obj by propWd_cns	0.62	.042	4.1 (1)
subj by propWd_cns	1.24	.552	0.4 (1)
propWd_cl	4.24	.000	77.3 (1)
head_cns by propSyl_ph	2.76	.016	5.8 (1)
constit_type * prom_wd	-	.000	20.0 (3)
adjunct by prom_wd	1.38	.009	6.8 (1)
obj by prom_wd	1.16	.201	1.6 (1)
subj by prom_wd	1.05	.789	0.1 (1)
POS_gp * prom_wd	-	.000	121.4 (6)
RB by prom_wd	2.16	.000	26.0 (1)
JJ by prom_wd	4.29	.000	82.4 (1)
PR by prom_wd	1.90	.000	14.3 (1)
VB by prom_wd	2.85	.000	47.1 (1)
NN by prom_wd	3.13	.000	76.4 (1)
DT by prom_wd	1.90	.000	15.8 (1)
numSyl_wd	1.45	.000	42.9 (1)
Constant	0.08	.000	76.3 (1)

F.6 Multivariate Test Results for K2

Table F.8: Multivariate Tests of All Factors in MANCOVA predicting acoustic features by accent status (Pillai's, Hotelling's, Wilk's)

Factor	F	df	Sig
propSyl_ph	93.5	4,4981	.000
numWd_ph	7.6	4,4981	.000
accPh_inc	34.3	4,4981	.000
npmean_ph	1602	4,4981	.000
nimean_ph	1036	4,4981	.000
acc_stat	30.2	8,9962	.000

Table F.9: Multivariate Tests of All Factors in MANCOVA predicting acoustic features of pre-nuclear accents (Pillai's, Hotelling's, Wilk's)

Factor	F	df	Sig
propSyl_ph	18.0	2,1918	.000
numWd_ph	9.5	2,1918	.000
accPh_inc	14.6	2,1918	.000
npmean_ph	939.4	2,1918	.000
nimean_ph	21.5	2,1918	.000
kon_stat	14.5	4,3836	.000

Table F.10: Multivariate Tests of All Factors in MANCOVA predicting acoustic features of nuclear accents (Pillai's, Hotelling's, Wilk's)

Factor	F	df	Sig
propSyl_ph	61.7	3,2817	.000
numWd_ph	2.8	3,2817	.038
accPh_inc	22.4	3,2817	.000
npmean_ph	1291	3,2817	.000
nimean_ph	13.8	3,2817	.000
kon_stat	8.4	6,5634	.000

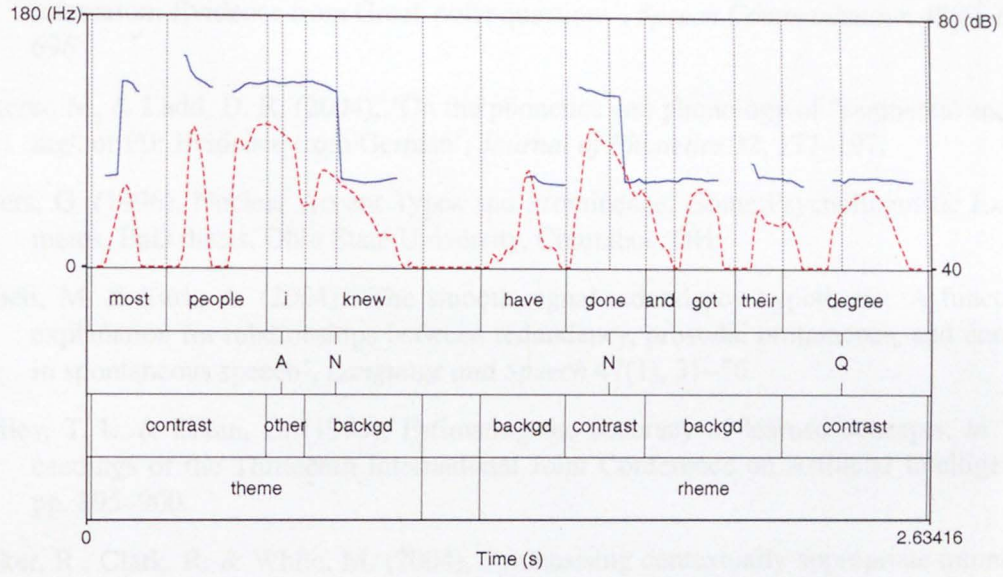
Table F.11: Multivariate Tests of All Factors in MANCOVA predicting peak features of nuclear accents (Pillai's, Hotelling's, Wilk's)

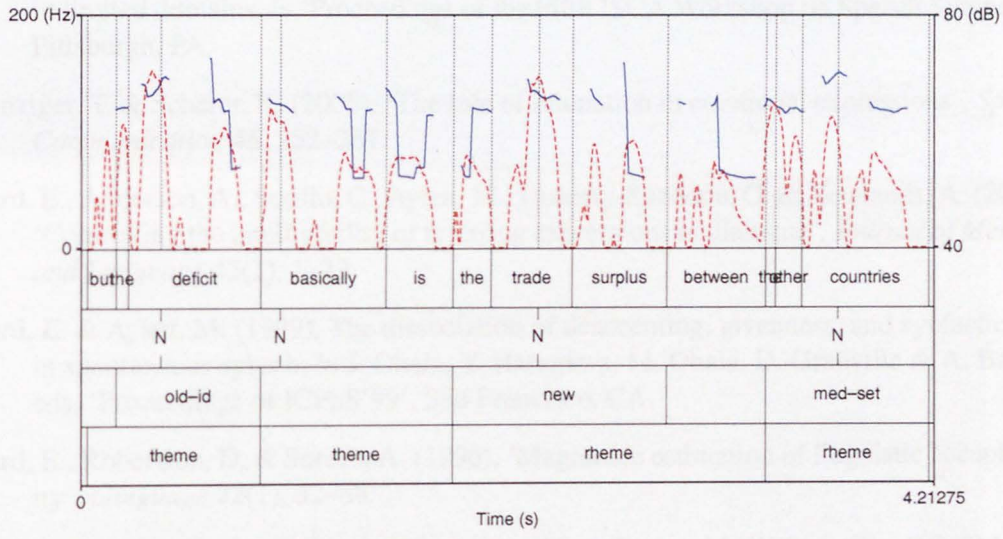
Factor	F	df	Sig
propSyl_ph	26.2	2,1986	.000
numWd_ph	4.7	2,1986	.009
accPh_inc	26.8	2,1986	.000
npmean_ph	734.9	2,1986	.000
kon_stat	8.1	4,3972	.000

Appendix G

Further Examples from Chapter 7

Pitch traces (blue line) and intensity curves (dashed red line) for (7.7) and (7.25).





Bibliography

Ariel, M. (1990), *Accessing Noun-Phrase Antecedents*, Routledge, London.

Arvaniti, A., Ladd, D. R. & Mennen, I. (1998), 'Stability of tonal alignment: the case of Greek prenuclear accents', *Journal of Phonetics* **26**, 3–25.

Arvaniti, A., Ladd, D. R. & Mennen, I. (2006), 'Effects of focus and "tonal crowding" in intonation: Evidence from Greek polar questions', *Speech Communication* **48**(6), 667–696.

Atterer, M. & Ladd, D. R. (2004), 'On the phonetics and phonology of "segmental anchoring" of F0: Evidence from German', *Journal of Phonetics* **32**, 177–197.

Ayers, G. (1996), Nuclear Accent Types and Prominence: Some Psycholinguistic Experiments, PhD thesis, Ohio State University, Columbus, OH.

Aylett, M. & Turk, A. (2004), 'The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech', *Language and Speech* **47**(1), 31–56.

Bailey, T. L. & Elkan, C. (1993), Estimating the accuracy of learned concepts, *in* 'Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence', pp. 895–900.

Baker, R., Clark, R. & White, M. (2004), Synthesising contextually appropriate intonation in limited domains, *in* 'Proceedings of the Fifth ISCA Workshop on Speech Synthesis', Pittsburgh, PA.

Bänziger, T. & Scherer, K. (2005), 'The role of intonation in emotional expressions', *Speech Communication* **46**, 252–267.

Bard, E., Anderson, A., Sotillo, C., Aylett, M., Doherty-Sneddon, G. & Newlands, A. (2000), 'Controlling the intelligibility of referring expressions in dialogue', *Journal of Memory and Language* **42**(1), 1–22.

Bard, E. & Aylett, M. (1999), The dissociation of deaccenting, givenness, and syntactic role in spontaneous speech, *in* J. Ohala, Y. Hasegawa, M. Ohala, D. Granville & A. Bailey, eds, 'Proceedings of ICPhS'99', San Francisco, CA.

Bard, E., Robertson, D. & Sorace, A. (1996), 'Magnitude estimation of linguistic acceptability', *Language* **72**(1), 32–68.

Bard, E., Sotillo, C., Anderson, A., Thompson, H. & Taylor, M. (1996), 'The DCIEM map task corpus: Spontaneous dialogue under sleep deprivation and drug treatment', *Speech Communication* **20**, 71–84.

Bartels, C. & Kingston, J. (1994), Salient pitch cues in the perception of contrastive focus, *in* R. van der Sandt, ed., 'Focus and natural language processing', Vol. 1, IBM Working Papers, Heidelberg, pp. 1–10.

Bates, E. & Devescovi, A. (1989), Crosslinguistic studies of sentence production, *in* B. MacWhinney & E. Bates, eds, 'The crosslinguistic study of sentence processing', Cambridge University Press, Cambridge, UK, pp. 225–256.

Bates, E. & MacWhinney, B. (1989), Functionalism and the competition model, *in* B. MacWhinney & E. Bates, eds, 'The crosslinguistic study of sentence processing', Cambridge University Press, Cambridge, UK, pp. 3–76.

Baumann, S. (2005), The Intonation of Givenness: Evidence from German, PhD thesis, Saarland University.

Baumann, S. & Grice, M. (2006), 'The intonation of accessibility', *Journal of Pragmatics* **38**(10), 1636–1657. Special Issue 'Prosody and Pragmatics'.

Beaver, D. & Clark, B. (2002), Monotonicity and focus sensitivity, *in* B. Jackson, ed., 'Proceedings of SALT XII', CLC Publications.

Beaver, D. & Clark, B. (2003), 'Always and only: Why not all focus sensitive operators are alike', *Natural Language Semantics* **11**(4), 323–362.

Beaver, D., Clark, B., Flemming, E., Jaeger, T. F. & Wolters, M. (2004), When semantics meets phonetics: Acoustical studies of second occurrence focus. Under submission.

Beckman, M. (1996), 'The parsing of prosody', *Language and Cognitive Processes* **11**(1/2), 17–67.

Beckman, M. & Elam, G. A. (1997), *Guidelines for ToBI Labelling*, The Ohio State University Research Foundation. version 3.0.

Beckman, M. & Hirschberg, J. (1999), 'The ToBI annotation conventions', `http://www.ling.ohio-state.edu/~tobi/ame_tobi/annotation_conventions.html`.

Beckman, M., Hirschberg, J. & Shattuck-Hufnagel, S. (2005), The original ToBI system and the evolution of the ToBI framework, *in* S.-A. Jun, ed., 'Prosodic models and transcription: Towards prosodic typology', Oxford University Press, Oxford, chapter 2.

Beckman, M. & Pierrehumbert, J. (1986), 'Intonational structure in English and Japanese', *Phonology Yearbook* **3**, 255–310.

Bell, A., Brenier, J. M., Gregory, M., Jurafsky, D. & Girand, C. (2004), Ranges and levels of predictability effects on word durations in conversational English, *in* 'Ninth Conference on Laboratory Phonology', University of Illinois at Urbana-Champaign.

Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., Gregory, M. & Gildea, D. (2003), 'Effects of disfluencies, predictability, and utterance position on word form variation in English conversation', *Journal of the Acoustical Society of America* **113**(2), 1001–1024.

Bies, A., Ferguson, M. & MacIntyre, R. (1995), 'Bracketing guidelines for Treebank II style', Ms., Department of Computer and Information Science, University of Pennsylvania.

Boersma, P. & Weenink, D. (2003), 'Praat:doing phonetics by computer', `http://www.praat.org`.

Böhmová, A., Hajič, J., Hajičová, E. & Hladká, B. (2001), The Prague dependency treebank: Three-level annotation scenario, *in* A. Abeillé, ed., 'Treebanks: Building and Using Syntactically Annotated Corpora', Kluwer Academic Publishers.

Bolinger, D. (1961), 'Contrastive accent and contrastive stress', *Language* **37**, 83–96.

Bolinger, D. (1965), Pitch accent and sentence rhythm, *in* I. Abe & T. Kanekiyo, eds, 'Forms of English: Accent, Morpheme and Order', Harvard University Press, Cambridge, MA, pp. 139–180.

Bolinger, D. (1972), 'Accent is predictable (if you're a mind reader)', *Language* **49**, 633–644.

Bolinger, D. (1978), Intonation across languages, *in* J. Greenberg, ed., 'Universals of Human Language. Volume II: Phonology', Stanford University Press, Palo Alto, CA, pp. 471–524.

Braun, B. (2005), Production and Perception of Contrastive and Non-Contrastive Themes in German, PhD thesis, Saarland University.

Brazil, D. (1975), *Discourse Intonation I*, Discourse Analysis Monographs, University of Birmingham.

Brazil, D. (1978), *Discourse Intonation II*, Discourse Analysis Monographs, University of Birmingham.

Brazil, D. (1985), *The Communicative Value of Intonation in English*, University of Birmingham. Second Edition, 1997, Cambridge University Press.

Breiman, L., Friedman, J., Olshen, R. & Stone, C. (1984), *Classification and Regression Trees*, Wadsworth, Belmont, CA.

Brenier, J. M., Cer, D. M. & Jurafsky, D. (2005), Emphasis detection in speech using acoustic and lexical features, *in* 'Annual Meeting of the Linguistics Society of America', Oakland, CA.

Brown, G., Currie, K. & Kenworthy, J. (1980), *Questions of Intonation*, Croom Helm, London.

Buráňová, E., Hajičová, E. & Sgall, P. (2000), Tagging of Very Large Corpora: Topic-Focus Articulation, *in* 'Proceedings of COLING Conference', Saarbrücken, Germany, pp. 278–284.

Büring, D. (2003), 'On D-trees, beans and B-accents', *Linguistics and Philosophy* **26**(5), 511–545.

Büring, D. (2004), 'Focus suppositions', *Theoretical Linguistics* **30**(1), 65–76.

Büring, D. (submitted), Intonation, semantics and information structure, *in* G. Ramchand & C. Reiss, eds, 'Interfaces'.

Büring, D. (to appear), Focus projection and default prominence, *in* V. Molnar & S. Winkler, eds, 'Proceedings of the Symposion Informationsstrukter-Kontrastivt'.

Byrd, S. & Clifton, C. (1995), 'Focus, accent and argument structure: Effects on language comprehension', *Language and Speech* **38**, 365–391.

Calhoun, S., Nissim, M., Steedman, M. & Brenier, J. (2005), A framework for annotating information structure in discourse, *in* 'Frontiers in Corpus Annotation II: Pie in the Sky, ACL2005 Conference Workshop', Ann Arbor, Michigan.

Cambier-Langeveld, T. & Turk, A. (1999), 'A cross-linguistic study of accentual lengthening: Dutch v English', *Journal of Phonetics* **27**, 171–206.

Campbell, N. & Beckman, M. (1997), Stress, prominence and spectral tilt, *in* A. Botinis, G. Kouroupetroglou & G. Carayiannis, eds, 'Intonation: Theory, Models and Applications (Proceedings of an ESCA Workshop)', Athens, Greece.

Carletta, J. (1996), 'Assessing agreement on classification tasks: the kappa statistic', *Computational Linguistics* **22**(2), 249–254.

Carletta, J., Dingare, S., Nissim, M. & Nikitina, T. (2004), Using the NITE XML toolkit on the Switchboard corpus to study syntactic choice: a case study, *in* 'Proceedings of LREC2004', Lisbon, Portugal.

Carletta, J., Evert, S., Heid, U. & Kilgour, J. (in press), 'The NITE XML Toolkit: data model and query language', *Language Resources and Evaluation Journal*.

Carletta, J., Evert, S., Heid, U., Kilgour, J., Robertson, J. & Voormann, H. (2003), 'The NITE XML Toolkit: flexible annotation for multi-modal language data', *Behavior Research Methods, Instruments and Computers* **35**(3), 353–363.

Carlson, R. & Granström, B. (1986), 'A search for durational rules in a real-speech database', *Phonetica* **43**, 140–154.

Chavarría, S., Yoon, T.-J., Cole, J. & Hasegawa-Johnson, M. (2004), Acoustic differentiation of ip and IP boundary levels: Comparison of L- and L-L% in the Switchboard corpus, *in* 'Speech Prosody 2004', Nara, Japan.

Chen, K. & Hasegawa-Johnson, M. (2004), An automatic prosody labeling system using ANN-based syntactic-prosodic model and GMM-based acoustic-prosodic model, *in* 'Proc. of ICASSP'.

Chomsky, N. (1957), *Syntactic Structures*, Mouton, The Hague.

Chomsky, N. & Halle, M. (1968), *The sound pattern of English*, Harper & Row, New York.

Chorianopoulou, E. (2002), Evaluating prodosy prediction in synthesis with respect to Modern Greek prenuclear accents, Master's thesis, Linguistics, University of Edinburgh.

Church, K. & Hanks, P. (1989), Word association norms, mutual information and lexicography, *in* 'Proceedings of ACL'.

Clark, H. (1996), *Using Language*, Cambridge University Press, Cambridge, UK.

Clark, R. (2003), Generating Synthetic Pitch Contours Using Prosodic Structure, PhD thesis, University of Edinburgh, Edinburgh, UK.

Clark, R. & King, S. (2006), Joint prosodic and segmental unit selection speech synthesis, *in* 'Interspeech', Pittsburgh, PA.

Cohen, A. & 't Hart, J. (1967), 'On the anatomy of intonation', *Lingua* **19**, 177–192.

Conkie, A., Riccardi, G. & Rose, R. (1999), Prosody recognition from speech utterances using acoustic and linguistic based models of prosodic events, *in* 'Proc. of Eurospeech', Budapest, Hungary, pp. 523–526.

Cooper, W. & Paccia-Cooper, J. (1980), *Syntax and Speech*, Harvard University Press, Cambridge, MA.

Core, M. & Allen, J. (1997), Coding dialogs with the DAMSL annotation scheme, *in* 'AAAI Fall Symposium on Communicative Action in Humans and Machines', Cambridge, MA.

Crystal, D. (1969), *Prosodic systems and intonation in English*, Cambridge University Press, Cambridge, UK.

Cutler, A. (1977), The context-dependence of "intonational meanings", *in* 'Papers from the 13th Meeting of the Chicago Linguistics Society', Chicago, IL, pp. 104–115.

Cutler, A., Dahan, D. & van Donselaar, W. (1997), 'Prosody in the comprehension of spoken language: A literature review', *Language and Speech* **40**(2), 141–201.

Dainora, A. (2001), An empirically based probabilistic model of intonation in American English, PhD thesis, University of Chicago.

Dainora, A. (2002), Does intonational meaning come from tones or tunes? Evidence against a compositional approach, *in* 'Speech Prosody 2002', Aix-en-Provence.

de Pijper, J. & Sanderman, A. (1994), 'On the perceptual strength of prosodic boundaries and its relation to suprasegmental cues', *Journal of the Acoustical Society of America* **96**(4), 2037–2047.

Deshmukh, N., Ganapathiraju, A., Gleeson, A., Hamaker, J. & Picone, J. (1998), Resegmentation of Switchboard, *in* 'Proceedings of ICSLP', Sydney, Australia, pp. 1543–1546.

Dilley, L. (2005), The phonetics and phonology of tonal systems, PhD thesis, MIT.

Dilley, L. & Brown, M. (2005), *The RaP (Rhythm and Pitch) Labeling System*, 1.0 edn, MIT and Ohio State University.

Eady, S. & Cooper, W. (1986), 'Speech intonation and focus location in matched statements and questions', *Journal of the Acoustical Society of America* **80**(2), 402–415.

Eady, S., Cooper, W., Klouda, W., Mueller, G. & Lotts, D. (1986), 'Acoustical characteristics of sentential focus: Narrow vs. broad and single vs. dual focus environments', *Language and Speech* **29**, 233–251.

Eckert, M. & Strube, M. (2001), 'Dialogue acts, synchronising units and anaphora resolution', *Journal of Semantics* **17**(1), 51–89.

Entropic-Research-Labs (1998), '*xwaves* manual'. version 5.3.1.

Fear, B., Cutler, A. & Butterfield, S. (1995), 'The strong/weak syllable distinction in English', *Journal of the Acoustical Soceity of America* **97**(3), 1893–1904.

Féry, C. (1993), *German Intonation Patterns*, Niemeyer, Tübingen.

Féry, C. & Samek-Lodovici (2006), 'Focus projection and prosodic prominence in nested foci', *Language* **82**(1), 131–150.

Fujisaki, H. (1981), Dynamic characteristics of voice fundamental frequency in speech and singing, *in* 'Presented at the 4th F.A.S.E. Symposium', Venice.

Fujisaki, H. (1996), Prosody, models and spontaneous speech, *in* Y. Sagisaka, N. Campbell & N. Higuchi, eds, 'Computing Prosody', Springer, Verlag, pp. 27–42.

Gee, J. & Grosjean, F. (1983), 'Performance structures: A psycholinguistic and linguistic appraisal', *Cognitive Psychology* **15**, 411–458.

German, J., Pierrehumbert, J. & Kaufmann, S. (2006), 'Evidence for phonological constraints on nuclear accent placement', *Language* **82**(1), 151–168.

Gerrits, E. (2001), The categorisation of speech sounds by adults and children. A study of the categorical perception hypothesis and the developmental weighting of acoustic speech cues, PhD thesis, Utrecht University, The Netherlands. LOT series, 42.

Gerrits, E. & Schouten, B. (1998), Categorical perception of vowels, *in* 'Proceedings of ICSLP'98', Sydney, Australia.

Godfrey, J., Holliman, E. & McDaniel, J. (1992), SWITCHBOARD: Telephone speech corpus for research and development, *in* 'Proceedings of ICASSP-92', pp. 517–520.

Grabe, E., Gussenhoven, C., Haan, J., Marsi, E. & Post, B. (1998), 'Preaccentual pitch and speaker attitude in Dutch', *Language and Speech* **41**(1), 63–85.

Grabe, E. & Low, E. (2002), Acoustic correlates of rhythm classes, *in* C. Gussenhoven & N. Warner, eds, 'Papers in Laboratory Phonology VII', Mouton de Gruyter, Berlin.

Grabe, E. & Warren, P. (1995), Stress shift: do speakers do it or do listeners hear it?, *in* B. Connell & A. Arvaniti, eds, 'Papers in Laboratory Phonology IV', Cambridge University Press, Cambridge, UK, pp. 95–110.

Greenberg, S., Hollenback, J. & Ellis, D. (1996), Insights into spoken language gleaned from phonetic transcription of the Switchboard corpus, *in* 'ICSLP', Philadelphia, PA.

Grice, M., Baumann, S. & Benzmüller, R. (2005), German intonation in autosegmental-metrical phonology, *in* S.-A. Jun, ed., 'Prosodic Typology', Oxford University Press, Oxford, pp. 53–83.

Grice, M., Ladd, D. R. & Arvaniti, A. (2000), 'On the place of phrase accents in intonational phonology', *Phonology* **17**(2), 143–185.

Grosjean, F., Grosjean, L. & Lane, H. (1979), 'The patterns of silence: Performance structures in sentence production', *Cognitive Psychology* **11**, 58–81.

Grosz, B. & Hirschberg, J. (1992), Some intonational characteristics of discourse structure, *in* 'Proceedings of the International Conference on Spoken Language Processing', Banff, Canada, pp. 429–432.

Grosz, B., Joshi, A. & Weinstein, S. (1983), Providing a united account of the definite noun phrases in discourse, *in* 'Proceedings of the 21st ACL', pp. 44–50.

Grosz, B. & Sidner, C. (1986), 'Attention, intentions, and the structure of discourse', *Computational Linguistics* **12**, 175–204.

Grosz, Barbara, A. J. & Weinstein, S. (1995), 'Centering: A framework for modelling the local coherence of discourse', *Computational Linguistics* **21**(2), 203–225.

Gundel, J. (1985), 'Shared knowledge and topicality', *Journal of Pragmatics* **9**(1), 83–107.

Gundel, J., Hedberg, N. & Zacharaski, R. (1993), 'Cognitive status and the form of referring expressions', *Language* **69**(2), 274–307.

Gunlogson, C. (2003), True to Form: Rising and Falling Declaratives as Questions in English, PhD thesis, UCSC.

Gussenhoven, C. (1983), 'Testing the reality of focus domains', *Language and Speech* **26**, 61–80.

Gussenhoven, C. (1984), *On the Grammar and Semantics of Sentence Accents*, Foris Publications, Dordrecht, chapter A semantic analysis of the nuclear tones of English.

Gussenhoven, C. (1988), Intonational phrasing and the prosodic hierarchy, *in* W. Dressler, H. Luschutsky, O. Pfeiffer & J. Rennison, eds, 'Phonologica', Cambridge University Press, Cambridge, UK, pp. 89–99.

Gussenhoven, C. (1999*a*), 'Discreteness and gradience in intonational contrasts', *Language and Speech* **42**(2-3), 283–305.

Gussenhoven, C. (1999*b*), On the limits of focus projection in English, *in* P. Bosch & R. van der Sandt, eds, 'Focus: Linguistic, Cognitive and Computational Perspectives', Cambridge University Press, Cambridge, UK, pp. 43–55.

Gussenhoven, C. (2002), Intonation and interpretation: phonetics and phonology, *in* B. Bel & I. Marlien, eds, 'Speech Prosody 2002', Aix-en-Provence, pp. 47–57.

Gussenhoven, C. (to appear), Types of focus in English, *in* D. Büring, M. Gordon & C. Lee, eds, 'Topic and Focus: Intonation and Meaning, Theoretical and Cross-Linguistic Perspectives', Kluwer, Dordrecht.

Gussenhoven, C. & Rietveld, A. (1988), 'Fundamental frequency declination in Dutch: Testing three hypotheses', *Journal of Phonetics* **16**, 355–69.

Hajičová, E. & Sgall, P. (2004), Degrees of contrast and the topic-focus articulation, *in* 'Information Structure: Theoretical and Empirical Aspects', Vol. 1 of *Language, Context and Cognition*, Mouton de Gruyter, Berlin, pp. 1–13.

Halle, M. & Vergnaud, J.-R. (1987), *An Essay on Stress*, MIT Press, Cambridge, MA.

Halliday, M. (1967), *Intonation and Grammar in British English*, The Hague, Mouton.

Halliday, M. (1968), 'Notes on transitivity and theme in English: Part 3', *Journal of Linguistics* **4**, 179–215.

Halliday, M. (1970), *A Course in Spoken English: Intonation*, Oxford University Press, UK.

Harkins, D. (2003), 'Switchboard resegmentation project', `http://www.cavs.msstate.edu/hse/ies/projects/switchboard`.

Harnad, S., ed. (1987), *Categorical perception: the groundwork of cognition*, Cambridge University Press, Cambridge, UK.

Harrington, J., Beckman, M., Fletcher, J. & Palethorp, S. (1998), An electropalatographic, kinematic, and acoustic analysis of supralaryngeal correlates of word and utterance-level prominence contrasts in English, *in* 'Proceedings of ICSLP'98', Sydney, Australia, pp. 1851–1854.

Haspelmath, M. (2006), 'Against markedness (and what to replace it with)', *Journal of Linguistics* **42**, 25–70.

Hayes, B. (1984), 'The phonology of rhythm in English', *Linguistic Inquiry* **15**, 33–74.

Hayes, B. (1989), The prosodic hierarchy in meter, *in* P. Kiparsky & G. Youmans, eds, 'Phonetics and Phonology, Vol I: Rhythm and Meter', Academic Press, San Diego, pp. 201–260.

Hayes, B. (1995), *Metrical Stress Theory*, University of Chicago Press, Chicago, IL.

Hedberg, N. & Sosa, J. M. (2001), The prosodic structure of topic and focus in spontaneous English dialogue, *in* 'Topic & Focus: A Workshop on Intonation and Meaning', LSA Summer Institute, University of California, Santa Barbara.

Hermes, D. & Rump, H. (1994), 'Perception of prominence in speech intonation induced by rising and falling pitch movements', *Journal of the Acoustical Society of America* **90**, 97–102.

Hirschberg, J. (1993), 'Pitch accent in context: Predicting intonational prominence from text', *Artificial Intelligence* **63**, 305–340.

Hirschberg, J. (2002), 'Communication and prosody: Functional aspects of prosody', *Speech Communication* **36**, 31–43.

Hirschberg, J. & Grosz, B. (1992), Intonational features of local and global discourse structure, *in* 'Proceedings of the Speech and Natural Language Workshop', DARPA, Harriman, NY, pp. 441–446.

Hirschberg, J. & Ward, G. (1992), 'The influence of pitch range, duration, amplitude and spectral features on the interpretation of the rise-fall-rise intonation contour in English', *Journal of Phonetics* **20**, 241–251.

Hirst, D. & Cristo, A. D. (1999), A survey of intonation systems, *in* D. Hirst & A. D. Cristo, eds, 'Intonation Systems', Cambridge University Press, Cambridge, UK, pp. 1–44.

Horne, M. (1988), 'Towards a quantified, focus-based model for synthesizing English sentence intonation', *Lingua* **75**, 25–54.

Hume, E. (2004), Deconstructing markeness: A predictability-based approach, *in* 'Proceedings of the Berkeley Linguistic Society'.

Huss, V. (1978), 'English word stress in the postnuclear position', *Phonetica* **35**, 86–105.

Ito, K. & Speer, S. (2005), The effect of intonation on visual search: An eye-tracking study, *in* 'Experimental Pragmatics: Exploring the Cognitive Basis of Conversation', The British Academy, Cambridge, UK.

Ito, K., Speer, S. & Beckman, M. (2004), Informational status and pitch accent distribution in spontaneous dialogues in English, *in* 'Speech Prosody 2004', Nara, Japan.

Jackendoff, R. (1995), The boundaries of the lexicon, *in* M. Everaert, E.-J. V. Linden & R. Schreuder, eds, 'Idioms, Structural and psychological perspectives', Erlbaum, Hillsdale, NJ, pp. 133–165.

Jackendoff, R. S. (1972), *Semantic Interpretation in Generative Grammar*, MIT Press, Cambridge, MA.

Jaeger, T. F. & Wagner, M. (2003), When warriors mourn longer. An investigation of some phonetic predictions of current semantic focus theories, *in* 'Semantics Fest 2003', Stanford University, CA.

Jakobson, R. (1963), *Roman Jakobson: Selected Writings II*, Mouton, The Hague, chapter Implications of language universals for linguistics.

Jakobson, R. & Pomorska, K. (1990), The concept of mark, *in* L. Waugh & M. Monville-Burston, eds, 'On Language: Roman Jakobson', Harvard University Press, Cambridge, MA, pp. 134–140.

Jurafsky, D. (2003), Probabilistic modeling in psycholinguistics: Linguistic comprehension and production, *in* H. Bod, J. Hay & S. Jannedy, eds, 'Probabilistic Linguistics', MIT Press, Cambridge, MA.

Jurafsky, D. & Martin, J. (2000), *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, Prentice Hall, Upper Saddle River, NJ.

Jurafsky, D., Shriberg, E. & Biasca, D. (1997), *Switchboard-DAMSL Labeling Project Coder's Manual*, tech. rep. no. 97-02 edn, University of Colorado Institute of Cognitive Science. `http://stripe.colorado.edu/~jurafsky/manual.august1.html`.

King, S. (2001), Speech and language processing. Course Notes, University of Edinburgh.

King, S., Black, A. W., Taylor, P., Caley, R. & Clark, R. (2003), *Edinburgh Speech Tools Library: System Documentation*, 1.2 edn, University of Edinburgh, Edinburgh.

Kiss, K. E. (1998), 'Identificational focus versus information focus', *Language* **74**(2), 245–273.

Kluender, K., Coady, J. & Kiefte, M. (2003), 'Sensitivity to change in perception of speech', *Speech Communication* **41**, 59–69.

Kochanski, G., Grabe, E., Coleman, J. & Rosner, B. (2005), 'Loudness predicts prominence: Fundamental frequency lends little', *Journal of the Acoustical Society of America* **118**(2), 1038–1054.

Krahmer, E. & Swerts, M. (2001), 'On the alleged existence of contrastive accents', *Speech Communication* **34**(4), 391–405.

Krifka, M. (1991), A compositional semantics for multiple focus constructions, *in* J. Jacobs, ed., 'Proceedings of Semantic and Linguistic Theory I', Cornell Working Papers in Linguistics 10, pp. 17–53.

Krifka, M. (2006), Association with focus phrases, *in* V. Molnar & S. Winkler, eds, 'The Architecture of Focus', Mouton de Gruyter, Berlin, New York.

Kruijff-Korbayová, I. & Steedman, M. (2003), 'Discourse and information structure', *Journal of Logic, Language and Information* **12**, 249–259.

Ladd, D. R. (1980), *The structure of intonational meaning: evidence from English*, Indiana University Press, Bloomington.

Ladd, D. R. (1983), 'Phonological features of intonational peaks', *Language* **59**, 721–759.

Ladd, D. R. (1986), 'Intonational phrasing: The case for recursive prosodic structure', *Phonology Yearbook* **3**, 311–340.

Ladd, D. R. (1988), 'Declination 'reset' and the hierarchical organization of utterances', *Journal of the Acoustical Society of America* **84**(2), 530–544.

Ladd, D. R. (1993), Constraints on the gradient variability of pitch range or pitch level 4 lives!, *in* P. Keating, ed., 'Papers in Laboratory Phonology 3', Cambridge University Press, Cambridge, UK, pp. 43–63.

Ladd, D. R. (1996), *Intonational Phonology*, Cambridge University Press, Cambridge, UK.

Ladd, D. R. (2004), Segmental anchoring of pitch movements: autosegmental phonology or speech production?, *in* H. Quené & V. van Heuven, eds, 'On Speech and Language: Essays for Sieb B. Nooteboom', LOT, Utrecht, pp. 123–131.

Ladd, D. R. & Campbell, N. (1991), Intonational phrasing: The case for recursive prosodic structure, *in* 'Proceedings of the XII International Congress of Phonetic Sciences', Aix-en-Provence, France.

Ladd, D. R., Faulkner, D., Faulkner, H. & Schepman, A. (1999), 'Constant "segmental anchoring" of F0 movements under changes in speech rate', *Journal of the Acoustical Society of America* **106**(3), 1543–1554.

Ladd, D. R., Mennen, I. & Schepman, A. (2000), 'Phonological conditioning of peak alignment of rising pitch accent in Dutch', *Journal of the Acoustical Society of America* **107**, 2685–2696.

Ladd, D. R. & Morton, R. (1997), 'The perception of intonational emphasis: continuous or categorical?', *Journal of Phonetics* **25**, 313–342.

Ladd, D. R. & Schepman, A. (2003), '"Sagging transitions" between high pitch accents in English: experimental evidence', *Journal of Phonetics* **31**(1), 81–112.

Ladd, D. R., Silverman, K., Tolkmitt, F., Bergmann, G. & Scherer, K. (1985), 'Evidence for the independent function of intonation contour type, voice quality and F0 range in signalling speaker affect', *Journal of the Acoustical Society of America* **78**, 435–444.

Ladd, D. R. & Terken, J. (1995), Modelling intra- and inter-speaker pitch range variation, *in* 'Proceedings of ICPhS', Vol. 2, Stockholm, pp. 386–9.

Ladd, D. R., Verhoeven, J. & Jacobs, K. (1994), 'Influence of adjacent pitch accents on each other's perceived prominence: two contradictory effects', *Journal of Phonetics* **22**, 87–99.

LDC (n.d.), Addendum to the pos tagging guidelines. `http://www.cis.upenn.edu/~bies/manuals/tagguid2.pdf`.

Levelt, W. (1989), *Speaking: From Intention to Articulation*, MIT Press, Cambridge, MA.

Liberman, M. (1975), The intonational system of English, PhD thesis, MIT Linguistics, Cambridge, MA.

Liberman, M. & Pierrehumbert, J. (1984), Intonational invariance under changes in pitch range and length, *in* M. Aronoff & R. Oerhle, eds, 'Language Sound Structure', MIT Press, Cambridge, MA, pp. 157–233.

MacDonald, M., N. P. & Seidenberg, M. (1994), The lexical nature of syntactic ambiguity resolution, *in* L. F. C. Clifton & K. Rayner, eds, 'Perspectives on sentence processing', Erlbaum, Hillsdale, NJ, pp. 123–154.

Mann, W. & Thompson, S. (1988), 'Rhetorical structure theory: Toward a functional theory of text organization', *Text* **8**(3), 243–281.

Marcus, M., Santorini, B. & Marcinkiewicz, M. A. (1993), 'Building a large annotated corpus of English: The Penn Treebank', *Computational Linguistics* **19**, 313–330.

Massaro, D. (1998), Categorical perception: Important phenomenon or lasting myth?, *in* 'Proceedings of ICSLP'98', Sydney, Australia.

Mateer, M. & Taylor, A. (1995), 'Disfluency annotation stylebook for the Switchboard corpus', Ms., Department of Computer and Information Science, University of Pennsylvania.

McNally, L. (1998), On recent formal analyses of topic, *in* J. Ginzburg, Z. Khasidashrili, C. Vogel, J. Lévy & E. Vallduví, eds, 'The Tbilisi Symposium on Language, Logic and Computation: Selected Papers', CLSI Publications, Stanford, CA, pp. 147–160.

Mücke, D. & Grice, M. (2005), Between prosodic structure and articulatory dynamics: the question of tonal alignment in German, *in* 'Phonetik und Phonologie 2', Universität Tübingen.

Nakatani, C. (1994), Discourse structural constraints on accent in spontaneous narrative, *in* 'Proceedings of the European Speech Communication Association/IEEE Workshop on Speech Synthesis', ESCA/IEEE, New Paltz, NY.

Nakatani, C., Hirschberg, J. & Grosz, B. (1995), Discourse structure in spoken language: Studies on speech corpora, *in* 'Working Notes of the AAAI Spring Symposium on Empirical Methods in Discourse Interpretation and Generation', Stanford, CA, pp. 106–112.

Nazzi, T. & Ramus, F. (2003), 'Perception and acquisition of linguistic rhythm by infants', *Speech Communication* **41**(1), 233–243.

Neeleman, A. & Szendröi, K. (2004), 'Superman sentences', *Linguistic Inquiry* **35**, 149–159.

Nespor, M. & Vogel, I. (1986), *Prosodic phonology*, Foris Publications, Dordrecht, Holland; Riverton, N.J.

Nissim, M. (2003), *Annotation Scheme for Information Status in Dialogue*, HCRC, University of Edinburgh. Internal Publication for Paraphrase project.

Nissim, M., Dingare, S., Carletta, J. & Steedman, M. (2004), An annotation scheme for information status in dialogue, *in* 'Fourth Language Resources and Evaluation Conference', Lisbon, Portugal.

O'Connor, J. & Arnold, G. (1961), *Intonation of Colloquial English*, Longmans, London.

Ohala, J. (1994), The frequency code underlines the sound symbolic use of voice of pitch, *in* L. Hinton, J. Nichols & J. Ohala, eds, 'Sound symbolism', Cambridge University Press, Cambridge, UK, pp. 325–247.

Oshima, D. (2002), Contrastive topic as paradigmatic operator, *in* 'Workshop on Information Structure in Context', Stuttgart University.

Ostendorf, M., Price, P. & Shattuck-Hufnagel, S. (1994), The Boston Unversity radio news corpus, Technical Report ECE-95-001, Boston University.

Ostendorf, M., Shafran, I., Shattuck-Hufnagel, S., Carmichael, L. & Byrne, W. (2001), A prosodically labeled database of spontaneous speech, *in* 'Proceedings of the ISCA Workshop on Prosody in Speech Recogniition and Understanding', Red Bank, NJ, pp. 119–121.

Pan, S. & McKeown, K. (1999), Word informativeness and automatic pitch accent modeling, *in* 'Proceedings of EMNLP/VLC99', College Park, Maryland.

Pan, S., McKeown, K. & Hirschberg, J. (2002), 'Exploring features from natural language generation for prosody modeling', *Computer Speech and Language* **16**, 457–490.

Partee, B. (1999), Focus, quantification, and semantics-pragmatics issues, *in* P. Bosch & R. van der Sandt, eds, 'Focus: Linguistic, Cognitive, and Computational Perspectives', Cambridge University Press, Cambridge, UK, pp. 213–231.

Pellom, B. (2001), SONIC: The University of Colorado continuous speech recognizer, Technical Report TR-CSLR-2001-01, University of Colorado at Boulder.

Pierrehumbert, J. (1980), The Phonology and Phonetics of English Intonation, PhD thesis, MIT, Cambridge, MA.

Pierrehumbert, J. (2000), Exemplar dynamics: Word frequency, lenition and contrast, *in* J. Bybee & P. Hopper, eds, 'Frequency effects and emergent grammar', John Benjamins, Amsterdam.

Pierrehumbert, J., Beckman, M. & Ladd, D. R. (2000), Conceptual foundations of phonology as a laboratory science, *in* N. Burton-Roberts, P. Carr & G. Docherty, eds, 'Phonological Knowledge: Its Nature and Status', Cambridge University Press, Cambridge, UK.

Pierrehumbert, J. & Hirschberg, J. (1990), The meaning of intonational contours in the interpretation of discourse, *in* P. Cohen, J. Morgan & M. Pollack, eds, 'Intentions in Communication', MIT Press, Cambridge, MA, pp. 271–311.

Pierrehumbert, J. & Steele, S. (1989), 'Categories of tonal alignment in English', *Phonetica* **46**, 181–196.

Pike, K. (1945), *The intonation of American English*, Unversity of Michigan Press, Ann Arbor, MI.

Pitrelli, J. (2004), ToBI prosodic analysis of a professional speaker of American English, *in* 'Speech Prosody 2004', Nara, Japan.

Pitrelli, J., Beckman, M. & Hirschberg, J. (1994), Evaluation of prosodic transcription labelling reliability in the ToBI framework, *in* 'Proceedings of the Third International Conference on Spoken Language Processing', Vol. 2, pp. 123–126.

Prevost, S. (1995), A Semantics of Contrast and Information Structure for Specifying Intonation in Spoken Language Generation, PhD thesis, University of Pennsylvania.

Prince, A. & Liberman, M. (1977), 'On stress and linguistic rhythm', *Linguistic Inquiry* **8**, 249–336.

Prince, E. (1981), Towards a taxonomy of given-new information, *in* P. Cole, ed., 'Radical Pragmatics', Academic Press, New York.

Prince, E. (1992), The ZPG letter: subjects, definiteness, and information-status, *in* S. Thompson & W. Mann, eds, 'Discourse Description: Diverse Analyses of a Fund Raising Text', John Benjamins, Philadelphia/Amsterdam, pp. 295–325.

Ramus, F., Nespor, M. & Mehler, J. (1999), 'Correlates of linguistic rhythm in the speech signal', *Cognition* **73**, 265–292.

Redi, L. (2003), Categorical effects in production of pitch contours in English, *in* 'Proceedings of the Fifteenth ICPhS', Barcelona, pp. 2921–2924.

Redi, L. & Shattuck-Hufnagel, S. (2001), 'Variation in the rate of glottalization in normal speakers', *Journal of Phonetics* **29**, 407–427.

Rietveld, A. & Gussenhoven, C. (1985), 'On the relation between pitch excursion size and prominence', *Journal of Phonetics* **13**, 299–308.

Roberts, C. (1998), 'Focus, the flow of information, and universal grammar', *Syntax and Semantics* **29**, 109–160.

Rochemont, M. (1986), *Focus in Generative Grammar*, John Benjamins, Philadelphia.

Rohde, D. (2005), *TGrep2 User Manual*, v 1.15 edn.

Roland, D. & Jurafsky, D. (2001), Verb sense and verb subcategorization probabilities, *in* P. Merlo & S. Stevenson, eds, 'Sentence processing and the lexicon: formal, computational, and experimental perspectives', Benjamins.

Rooth, M. (1992), 'A theory of focus intepretation', *Natural Language Semantics* **1**, 75–116.

Rooth, M. (1996*a*), Focus, *in* S. Lappin, ed., 'The Handbook of Contemporary Semantic Theory', Basil Blackwell, London, pp. 271–297.

Rooth, M. (1996*b*), On the interface principles for intonational focus, *in* 'Proceedings, Conference on Semantics and Linguistic Theory (SALT)', Vol. 6.

Rooth, M. (1999), Association with focus or association with presupposition, *in* P. Bosch & R. van der Sandt, eds, 'Focus: Linguistic, Cognitive and Computational Perspectives', Cambridge University Press, Cambridge, UK, pp. 232–244.

Rump, H. & Collier, R. (1996), 'Focus conditions and the prominence of pitch-accented syllables', *Language and Speech* **39**, 1–17.

Sag, I. & Liberman, M. (1975), The intonational disambiguation of indirect speech acts, *in* 'Proceedings of the Chicago Linguistics Society', Vol. 11, pp. 487–497.

Salton, G. (1989), *Automatic Text Processing: The Transformation, Analysis and Retrieval of Information by Computer*, Addison-Wesley, Reading, MA.

Santorini, B. (1990), Part-of-speech tagging guidelines for the Penn Treebank project, Technical Report MS-CIS-90-47, Department of Computer and Information Science, University of Pennsylvania.

Schafer, A. (1995), The role of optional prosodic boundaries, *in* 'Eighth Annual CUNY Conference on Human Sentence Processing', Tucson, Arizona.

Schafer, A., Carlson, K., Clifton, C. & Frazier, L. (2000), 'Focus and the interpretation of pitch accent: Disambiguating embedded questions', *Language and Speech* **43**(1), 75–105.

Schafer, A., Carter, J., Clifton, J. & Frazier, L. (1996), 'Focus in relative clause construal', *Language and Cognitive Processes* **11**, 135–163.

Schafer, A., Speer, S. & Warren, P. (2000), 'Intonational disambiguation in sentence production and comprehension', *Journal of Psycholinguistic Research* **29**(2), 169–182.

Schepman, A., Lickley, R. & Ladd, D. R. (2006), 'Effects of vowel length and "right context" on the alignment of Dutch nuclear accents', *Journal of Phonetics* **34**(1), 1–28.

Scherer, K. & Banziger, T. (2004), Emotional expression in prosody: A review and an agenda for future research, *in* 'Speech Prosody 2004', Nara, Japan.

Scherer, K., Ladd, D. R. & Silverman, K. (1984), 'Vocal cues to speaker affect: Testing two models', *Journal of the Acoustical Society of America* **76**(5), 1346–1356.

Schouten, B., Gerrits, E. & van Hessen, A. (2003), 'The end of categorical perception as we know it', *Speech Communication* **41**, 71–80.

Schwarzschild, R. (1999), 'Givenness, AVOIDF and other constraints on the placement of accent', *Natural Language Semantics* **7**, 141–177.

Selkirk, E. (1984), *Phonology and Syntax*, MIT Press, Cambridge, MA.

Selkirk, E. (1995), Sentence prosody: Intonation, stress and phrasing, *in* J. Goldsmith, ed., 'The Handbook of Phonological Theory', Blackwell, Cambridge, MA & Oxford, pp. 550–569.

Shannon, C. (1948), 'A mathematical theory of communication', *Bell System Technical Journal* **27**, 379–423;623–656.

Shattuck-Hufnagel, S., Dilley, L., Veilleux, N., Brugos, A. & Speer, R. (2004), F0 peaks and valleys aligned with non-prominence in adjacent syllables, *in* 'Speech Prosody 2004', Nara, Japan.

Shattuck-Hufnagel, S., Ostendorf, M. & Ross, K. (1994), 'Stress shift and early pitch accent placement in lexical items in American English', *Journal of Phonetics* **22**, 357–388.

Shattuck-Hufnagel, S. & Turk, A. E. (1996), 'A prosody tutorial for investigators of auditory sentence processing', *Journal of Psycholinguistic Research* **25**(2), 193–247.

Shriberg, E., Ladd, D. R., Terken, J. & Stolcke, A. (1996), Modeling pitch range variation within and across speakers: Predicting F0 targets when "speaking up", *in* 'Proceedings of ICSLP', Philadelphia, PA.

Shriberg, E., Stolcke, A., Hakkani-Tür, D. & Tür, G. (2000), 'Prosody-based automatic segmentation of speech into sentences and topics', *Speech Communication* **32**(2), 127–154.

Shriberg, E., Taylor, P., Bates, R., Stolcke, A., Ries, K., Jurafsky, D., Coccaro, N., Martin, R., Meteer, M. & Ess-Dykema, C. (1998), 'Can prosody aid the automatic classification of dialog acts in conversational speech?', *Language and Speech* **41**(3-4), 439–487.

Silverman, K., Beckman, M., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J. & Hirschberg, J. (1992), A standard for labelling English prosody, *in* 'Proceedings of the International Conference on Spoken Language Processing (ICSLP)', Vol. 2, Banff, pp. 867–870.

Silverman, K. & Pierrehumbert, J. (1990), The timing of prenuclear high accents in English, *in* J. Kingston & M. Beckman, eds, 'Papers in Laboratory Phonology I: Between the grammar and physics of speech', Cambridge University Press, Cambridge, UK, pp. 71–106.

Sluijter, A. & Terken, J. (1993), 'Beyond sentence prosody: Paragraph intonation in Dutch', *Phonetica* **50**, 180–188.

Sluijter, A. & van Heuven, V. (1996), 'Spectral balance as an acoustic correlate of linguistic stress', *Journal of the Acoustical Society of America* **100**(4), 2471–2485.

Sorace, A. & Keller, F. (2005), 'Gradience in linguistic data', *Lingua* **115**(1), 1497–1524.

Speer, S., Warren, P. & Schafer, A. (2003), Intonation and sentence processing, *in* 'Fifteenth International Congress of Phonetic Sciences', Barcelona.

Sprenger, S., Levelt, W. & Kempen, G. (2006), 'Lexical access during the production of idiomatic phrases', *Journal of Memory and Language* **54**, 161–184.

Stalnaker, R. C. (1978), Assertion, *in* P. Cole, ed., 'Syntax and Semantics', Vol. 9, Academic Press, New York, pp. 315–332.

Steedman, M. (2000), 'Information structure and the syntax-phonology interface', *Linguistic Inquiry* **31**(4), 649–689.

Steedman, M. (2001), *The Syntactic Process*, MIT Press, Cambridge,MA.

Steedman, M. (2006*a*), Information-structural semantics for English intonation, *in* C. Lee, M. Gordon & D. Büring, eds, 'LSA Summer Institute Workshop on Topic and Focus, Santa Barbara July 2001', Kluwer, Dordrecht, pp. 245–264.

Steedman, M. (2006*b*), 'Semantics and implicature in the meaning of English intonation', *submitted* .

Stone, M. (1998), Modality in Dialogue: Planning Pragmatics and Computation, PhD thesis, University of Pennsylvania.

Stone, M. (2004), 'Intention, interpretation, and the computational structure of language', *Cognitive Science* **28**, 781–809.

Strom, V., Clark, R. & King, S. (2006), Expressive prosody for unit-selection speech synthesis, *in* 'Proceedings of the International Conference on Spoken Language Processing', Pittsburgh, PA.

Suci, G. (1967), 'The validity of pause as an index of units in language', *Journal of Verbal Learning and Verbal Behaviour* **6**, 26–32.

Sugahara, M. (2003), Downtrends and Post-FOCUS Intonation in Tokyo Japanese, PhD thesis, University of Massachusetts, Amherst.

Swerts, M. (1997), 'Prosodic features at discourse boundaries of different strengths', *Journal of the Acoustical Society of America* **101**, 514–521.

Syrdal, A. & McGory, J. (2000), Inter-transcriber reliability of ToBI prosodic labeling, *in* 'ICSLP2000', Beijing, China.

't Hart, J. & Cohen, A. (1973), 'Intonation by rule: a perceptual quest', *Journal of Phonetics* **1**, 309–327.

't Hart, J. & Collier, R. (1975), 'Integrating different levels of intonation analysis', *Journal of Phonetics* **3**, 235–255.

Tabor, W., Cornell, J. & Tanenhaus, M. (1997), 'Parsing in a dynamical system', *Language and Cognitive Processes* **12**, 211–272.

Taylor, A. (1996), 'Bracketing Switchboard: An addendum to the Treebank II bracketing guidelines', `http://www.stanford.edu/dept/linguistics/corpora/BracketingSwitchboard.pdf`.

Taylor, A., Marcus, M. & Santorini, B. (2003), 'The Penn Treebank: An overview', `http://citeseer.ist.psu.edu/taylor03penn.html`.

Taylor, P. (2000), 'Analysis and synthesis of intonation using the Tilt model', *Journal of the Acoustical Society of America* **107**, 1697–1714.

Terken, J. (1991), 'Fundamental frequency and perceived prominence of accented syllables', *Journal of the Acoustical Society of America* **89**, 1768–1776.

Terken, J. & Hermes, D. (2000), The perception of prosodic prominence, *in* M. Horne, ed., 'Prosody: Theory and experiment. Studies presented to Gösta Bruce', Dordrecht: Kluwer, pp. 89–127.

Terken, J. & Hirschberg, J. (1994), 'Deaccentuation of words representing 'given' information: Effects of persistence of grammatical role and surface position', *Language and Speech* **37**, 125–145.

Truckenbrodt, H. (1995), Phonological Phrases: Their Relation to Syntax, Focus and Prominence, PhD thesis, MIT.

Truckenbrodt, H. (1999), 'On the relation between syntactic phrases and phonological phrases', *Linguistic Inquiry* **30**(2), 219–255.

Truckenbrodt, H. (2002), 'Upstep and embedded register levels', *Phonology* **19**, 77–120.

Truckenbrodt, H. (2006), Phrasal stress, *in* K. Brown, ed., 'The Encyclopedia of Languages and Linguistics', 2nd edn, Vol. 9, Elsevier, Oxford, UK.

Trueswell, J. (1996), 'The role of lexical frequency in syntactic ambiguity resolution', *Journal of Memory and Language* **35**, 566–585.

Trueswell, J., Tanenhaus, M. & Garnsey, S. (1994), 'Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution', *Journal of Memory and Language* **33**, 285–318.

Turk, A. (1999), 'Structural influences on accentual lengthening in English', *Journal of Phonetics* **27**, 171–206.

Turk, A. & Sawusch, J. (1996), 'The processing of duration and intensity cues to prominence', *Journal of the Acoustical Society of America* **99**(6), 3782–3790.

Turk, A. & Sawusch, J. (1997), 'The domain of accentual lengthening in American English', *Journal of Phonetics* **25**, 25–41.

Turk, A. & Shattuck-Hufnagel, S. (2000), 'Word-boundary-related duration patterns in English', *Journal of Phonetics* **28**, 397–440.

Umbach, C. (2004), 'On the notion of contrast in information structure and discourse structure', *Journal of Semantics* **21**, 155–175.

Vallduví, E. & Vilkuna, M. (1998), 'On rheme and kontrast', *Syntax and Semantics* **29**, 79–108.

van Santen, J. & Möbius, B. (2000), A quantitative model of F0 generation and alignment, *in* A. Botinis, ed., 'Intonation', Kluwer Academic Publishers, pp. 269–288.

van Wijk, C. (1987), 'The PSY behind PHI: A psycholinguistic model for performance structures', *Journal of Psycholinguistic Research* **16**(2), 185–199.

Veselá, K., Havelka, J. & Hajičová, E. (2004), Annotators' agreement: The case of topic-focus articulation, *in* 'Proceedings of the 4th International Conference on Language Resources and Evaluation', European Language Resources Association, pp. 2191–2194.

Šafářová, M. & Swerts, M. (2004), On recognition of declarative questions in English, *in* 'Speech Prosody 2004', Nara, Japan.

Wagner, M. (2003), Prosody as diagonalization of syntax. evidence for complex predicates, *in* K. Moulton & M. Wolf, eds, 'Proceedings of NELS34', SUNY Stony Brook.

Wagner, M. (2005), Asymmetries in prosodic domain formation, *in* N. Richards & M. McGinnis, eds, 'Perspectives on Phrases', MITWPL 49, Cambridge, MA, pp. 329–367.

Wagner, M. (2006), Givenness and locality, *in* M. Gibson & J. Howell, eds, 'Proceedings of SALT XVI'.

Ward, G. & Hirschberg, J. (1985), 'Implicating uncertainty: The pragmatics of fall-rise intonation', *Language* **61**, 747–776.

Ward, G. & Hirschberg, J. (1986), Reconciling uncertainty with incredulity: A unified account of the L*+H LH% intonational contour, *in* 'Linguistic Society of America Annual Meeting'.

Warren, P. (1999), Prosody and language processing, *in* S. Garrod & M. Pickering, eds, 'Language Processing', Pyschology Press, UK, pp. 155–188.

Watson, D. & Arnold, J. (2005), Not just given and new: the effects of discourse and task-based constraints on acoustic prominence, *in* 'CUNY 2005', Tucson, AZ.

Watson, D., Tanenhaus, M. & Gunlogson, C. (2004), Processing pitch accents: Interpreting H* and L+H*, *in* 'Presented at the 17th Annual CUNY Conference on Human Sentence Processing', Cambridge, MA.

Welby, P. (2003), 'Effects of pitch accent position, type and status on focus projection', *Language and Speech* **46**(1), 53–81.

Wichmann, A., House, J. & Rietveld, A. (2000), Discourse constraints of F0 peak timing in English, *in* A. Botinis, ed., 'Intonation', Kluwer Academic Publishers, The Netherlands, pp. 163–182.

Wightman, C., Shattuck-Hufnagel, S., Ostendorf, M. & Price, P. (1992), 'Segmental durations in the vicinity of prosodic phrase boundaries', *Journal of the Acoustical Society of America* **91**(3), 1707–1717.

Wright, H. & Taylor, P. (1997), Modelling intonational structure using Hidden Markov Models, *in* 'ESCA Workshop on Intonation: Theory, Models and Applications'.

Xu, Y. (1999), 'Effects of tone and focus on the formation and alignment of F0 contours', *Journal of Phonetics* **27**, 55–105.

Xu, Y. (2005), 'Speech melody as articulatorily implemented communicative functions', *Speech Communication* **46**, 220–251.

Xu, Y., Ching, X. X. & Xuejing, S. (2004), On the temporal domain of focus, *in* 'Speech Prosody 2004', Nara, Japan.

Xu, Y. & Xu, C. (2005), 'Phonetic realization of focus in English declarative intonation', *Journal of Phonetics* **33**, 159–197.

Yoon, T.-J., Chavarría, S., Cole, J. & Hasegawa-Johnson, M. (2004), Intertranscriber reliability of prosodic labeling on telephone conversation using ToBI, *in* 'Proceedings of ICSLP', Jeju, Korea.

Zhang, T., Hasegawa-Johnson, M. & Levinson, S. (2006), 'Extraction of pragmantic and semantic salience from spontaneous spoken English', *Speech Communication* **48**, 437–462.